ERIC Document Resume	**See Instruction Sheet**	**DATE OF RESUME**	**DATE RECEIVED**

A. PUBLICATION DATA

1. ERIC NUMBER D001200

2. ORIGINATOR CONTROL NO(S). N.A.

3. IDENTIFICATION NO. CR 929

4. AUTHOR(S): Charters, W. W., Jr.

5. TITLE OF DOCUMENT ABSTRACTED: Teacher Perceptions of Administrator Behavior

6. PUBLICATION TITLE: N.A.

7. EDITOR(S): N.A.

8. PLACE OF ORIGIN
ORG. Washington Univ.
ST. CITY, STATE St. Louis, Mo.
TELE. EXT.
INST/LOC CODE

9. PUBLISHER OR SOURCE Cooperative Research Program, Office of Edu., Dept. of HEW., Washington, DC

10. VOL.	11. PAGES	12. MO./YR.	13. TOTAL PP.
N.A.	N.A.	Jan. 64	214

DO NOT WRITE IN SPACE BELOW

MICROFILM COPY AVAILABLE
YES XX NO ☐

COPYRIGHT GRANTED
YES ☐ NO ☒

COPYRIGHT RELEASE FILED
YES ☐ NO ☐

TYPE OF REPORT
PROPOSAL ☐
IN-PROGRESS ☐
FINAL ☒
FOLLOW-UP ☐

OTHER ERIC REFERENCES:

ERIC SATELLITE CODE

B. INDEXING DATA

14. SUBJECT AREA

15. TYPE OF ENDEAVOR report

16. INDEXING TERMS (6-12)

superintendent leadership	research replications
leadership measurements	school systems
teacher attitudes	administrator attitudes

17. ABSTRACT (MAX. 250 WORDS)

Part I replicated research concerning teacher attitudes toward administrators. Previous findings indicated teachers in small school systems had a stronger sense of superintendent leadership than did teachers in large school systems. However, superintendents responded with opposite perceptions in self-analysis tests. The replication recommended that the findings be regarded as entirely fortuitous.

Part II indicated the invalidity of the Leadership Behavior Description Questionnaire because of inconsistency of results. It was recommended that the investigator assess carefully the opportunities for his respondents to observe the kinds of behavior upon which they will be asked to report.

ERIC No.

18. SOURCE OF ABSTRACT

C. FUNDING DATA

19. SUPPORTING ORGANIZATION

20. DIVISION OR DEPT.

21. CONTRACT NO.

PROJECT NO.

22. TOTAL COST OF PROJECT

23. DURATION OR PROJECT

U.S. Department of Health, Education, and Welfare Office of Education

4

ERIC Form 1

929

ED001200

TEACHER PERCEPTIONS of ADMINISTRATOR BEHAVIOR

Cooperative Research Project No. 929

W. W. Charters, Jr.
Washington University
St. Louis, Missouri

January, 1964

REPRINTED BY PERMISSION
U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN PRINTED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

Teacher Perceptions of Administrator Behavior

Cooperative Research Project No. 929

W. W. Charters, Jr.
Washington University
St. Louis, Missouri

January, 1964

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

Acknowledgements

The U. S. Office of Education's Cooperative Research Program, jointly with Washington University, funded the research project reported here. The author also owes a profound debt to Dr. Robert J. Schaefer, Director of the Graduate Institute of Education at the time the project was initiated and carried out, for his encouragement of serious and scholarly study of educational affairs, and to his faculty colleagues whose dedication to scholarship transformed the encouragement into an Institute norm. Two persons worked closely with the author throughout the investigation, Dr. Gerald K. Moeller and Dr. Roland Keene, the former in the capacity of research assistant. Dr. Ovid N. Hunter made all of the data from his prior study fully accessible. Especial appreciation must go to Dr. Andrew W. Halpin, only later a member of the Institute staff, who took seriously the investigator's concerns with Hunter's findings and who went out of his way to contribute additional information from studies of his own.

Each of the following contributed her unique and indispensible skills to the completion of the research: Mrs. Carol Pauk, Mrs. Margaret Kovar, Mrs. Jane Saenger, Miss Julia Finn, Mrs. Lee Moeller, Miss Janet Williford, Mrs. Jeneane Johanningmeier, and Mrs. Charlotte Charters.

Nearly one thousand anonymous teachers and administrators in the public schools of the Midwest, in the final analysis, made the study possible.

Table of Contents

Chapter page

PART I

LEADER BEHAVIOR AND SCHOOL SYSTEM SIZE: A REPLICATION

1 Descriptions of Superintendent Behavior: A Problem Posed 2

The Plan of Part I 7

Attempted Confirmation: Data from Halpin 8

Ohio Superintendents
Illinois high school principals

Re-analysis of the Hunter Data 14

Preparation of a scoring key
Results of the re-analysis

Plans for a Replication 21

2 Methods of the Replication Study 23

Methods in the 1958 and 1961 Studies 24

School systems and their size classes
Respondents and their selection
Data-taking methods
The Leader Behavior Description Questionnaire
Summary of method differences

Additional Details of Procedure: 1961 Study 39

Unusable returns
Attributes of non-respondents

Summary 44

Contents (Contd.)

Chapter page

3 Results of the Replication Study 48

- Comparisons of LBDQ Scores by System Size 49
- Differences in Adjusted Size Classes 56
- Same System Comparisons 58
  - Superintendent self-descriptions
  - Teacher descriptions
- Size Differences in the Early Returns 63
- Size Differences in the "Personal" Form of the LBDQ 65
- Summary 68

4 Conclusions of the Replication Study 71

PART II

THE LBDQ AND SUPERINTENDENT BEHAVIOR

5 Introduction 75

- The Plan of Part II 76

6 Measures of Teacher Affect and Superintendent Contact 78

- Teacher Contact with the Superintendent 78
  - Questionnaire items
  - The contact index
- Measures of Teacher Affect 83
  - Initial scale construction
  - Re-examination of the scales
  - Factor analysis of the scale items

Contents (Contd.)

Chapter page

7. The LBDQ as a Veridical Report on Superintendent Behavior 97

Relationship with Other Variables 99

Discriminations in the Behavior of a Leader 105

Independence of dimensions
Halo in the LBDQ
Halo in the "personal" form of the LBDQ

Discrimination between Several Leaders 118

Agreement among Respondents on a Given Leader 119

Between-group agreement
Within-group agreement

Summary 128

8 Reinterpretation of the LBDQ Scores 131

Two Determinants of LBDQ Responses 131

A Test of the Joint Determinants 133

9 Relevance of the LBDQ for the School Executive 138

Role of the Superintendent vis-a-vis Teachers 140

Intervening levels of authority
Remoteness of teachers from the superintendent
Difficulties in responding to the LBDQ

The Institutionalization of Leadership 155

The structure of organization
Professional norms
Consideration in bureaucratic organization

Summary 164

Contents (Cont.)

Chapter page

10 Summary of Parts I and II 168

The Preliminary Search for Confirmation 169

The Replication of Hunter's Investigation 172

Reconsideration of the LBDQ in School Systems 176

A Note on Descriptions of Leader Behavior 181

Leadership functions in the organization
The leader-follower relationship
Kinds of information about behavior
Access to information about behavior
Behavior versus reactions to behavior

Appendices 190

I Mean LBDQ Scores by Teaching Level and Size; Teachers (Hunter's Data) 190

II Items in "Personal" Form of the LBDQ 191

III LBDQ Data by School Systems: Replication Study 192

IV Items of the Teacher Affect Scales 194

V Scale Analysis Details: Trial Run Data 196

VI Factor Loadings of Items on Teacher Affect Scales: Trial Run Data 200

VII Teacher Reaction Questionnaire 201

List of Tables

page

1 Hunter's Data: Correlations of LBDQ Scores and Size; Superintendents, Teachers, and School Board 5

2 Halpin's Data: Mean LBDQ Scores by Size; Superintendents, Board Members, and Staff 11

3 Evenson's Data: Mean LBDQ Scores Attributed to Principals by High School Size; Superintendents, Principals, and Teachers 13

4 Hunter's LBDQ Items, Classified as Biased or Unbiased 19

5 Teachers' Scores on Biased and Unbiased LBDQ Items by Size 20

6 Size of School Systems in Hunter's Study and the Replication Study 27

7 Sex and Teaching Level of Responding and Nonresponding Teachers 45

8 Correlations of LBDQ Scores and Size; Superintendents and Teachers 50

9 Mean LBDQ Scores by Size; Superintendents and Teachers 51

10 Mean LBDQ Scores in 1958 and 1961 Studies by Size; Superintendents and Teachers 53

11 Correlations of LBDQ Scores and Size; Administrative Staff 55

12 Mean LBDQ Scores by Adjusted Size Categories; Superintendents and Teachers 57

13 LBDQ Scores from Same Superintendents in 1958 and 1961 60

14 LBDQ Scores Attributed to Same Superintendents by Teachers in 1958 and 1961 61

15 Mean LBDQ Scores by Size; Teachers Who Replied Early 66

16 Mean LBDQ Scores by Size; Teachers on "Personal" Form of the LBDQ 67

Tables (Cont.)

page

17 Intimacy by Frequency of Teacher Contact with Superintendent 81

18 Scale Analysis Details for Four Measures of Teacher Affect; Trial Run Data 88

19 Rescaling of Items in Four Measures of Teacher Affect; Study Population Data 90

20 Scale Score Distributions for Four Measures of Teacher Affect; Study Population Data 92

21 Cosines of Angles between Factors Corresponding to Five Measures of Teacher Affect 95

22 Correlations of Teacher LBDQ Responses with Two Measures of Teacher Morale 102

23 Correlations of Teacher LBDQ Responses, "Regular" Form, with Two Evaluations of Superintendents 110

24 Correlations of Teacher LBDQ Responses, "Personal" Form, with Two Evaluations of Superintendents 115

25 Correlations between Superintendents, Teachers, and Staff Responses on the LBDQ 124

26 Correlations of LBDQ Scores and Teacher Affect toward Superintendents; for Teachers with "Much" and "Little" Contact with Superintendent 135

27 Per Cent of Teachers with Varying Degrees of Intimacy and Frequency of Contact with Superintendent 145

28 Teacher Contacts with Superintendent in Large and Small School Systems 146

29 Estimate Number of Teachers in 20 School Systems with Intimate, Frequent Superintendent Contacts 148

30 Contact Scores by Degree of Personal Acquaintance with Superintendent in Large and Small School Systems 151

List of Figures

page

1 Hunter's Datq: LBDQ Scores by Size; Superintendents, Teachers, and School Board Members 4

2 Relationship between Rating Scale Evaluations and Scores on Three Initiating Structure Items 114

PART I

Leader Behavior and School System Size: A Replication

Chapter I

Descriptions of Superintendent Behavior

A Problem Posed

In the course of his dissertation research in 1958, O.N. Hunter[1] encountered a bizarre finding regarding the manner in which personnel in school systems of varying size described their superintendents. He had obtained responses on his adaptation of the Leader Behavior Description Questionnaire[2] from teachers, board members, and superintendents in 16 metropolitan area school systems, half of which were "small"--enrolling from 750 to 2700 pupils--and half of which were "large"--6,500 to 10,200 pupils. The Leader Behavior Description Questionnaire (LBDQ) is a thirty-item rating form providing two scores, one designed to measure a leader's organizing capacity (Initiating Structure) and the other his considerateness in dealing with subordinates (Consideration).

---

[1] Ovid N. Hunter, Relationship between School Size and Discrepancy in Perception of the Superintendent's Behavior, Unpublished doctoral dissertation, Washington University, 1959.

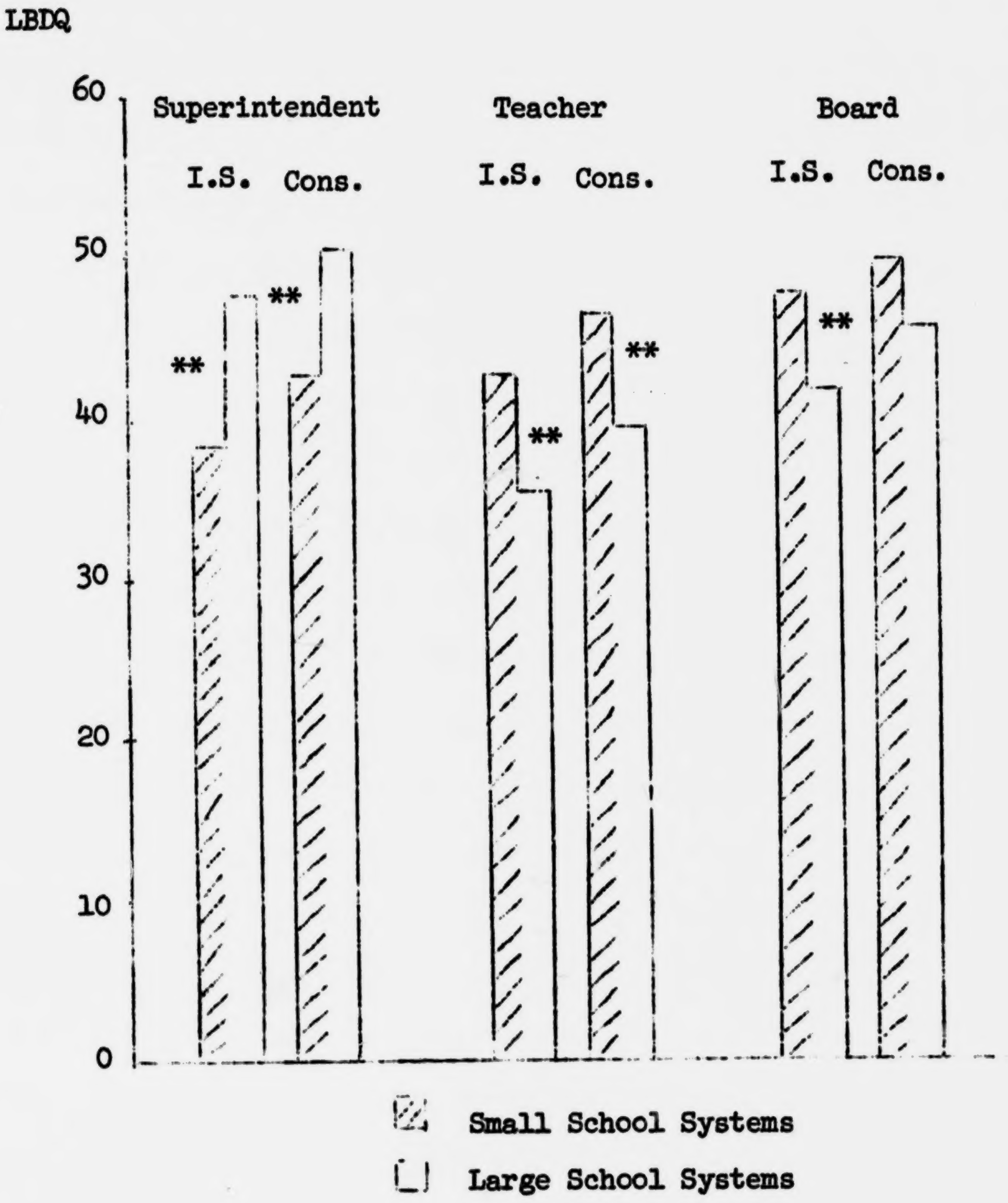
[2] Ralph M. Stogdill and Alvin E. Coons, (Eds.) Leader Behavior: Its Description and Measurement, Research Monograph No. 88, Columbus, Ohio: Bureau of Business Research, Ohio State University, 1957

To his surprise, Hunter observed that teachers and board members in the large school systems rated their superintendents lower on both dimensions of leadership than teachers and board members in the small systems. Just the reverse occurred in the superintendents' self ratings: superintendents in the large systems described themselves as higher in both Initiating Structure and Consideration than those in the small systems. (See Figure 1.) All but one of the differences were statistically significant (beyond the 1 per cent level) by the appropriate t test, and some of them were dramatically large in absolute terms. There was virtually no overlap in the distributions of scores in the case of the superintendents' self-ratings. The point biserial correlations between school system size and LBDQ scores shown in Table 1 indicate the magnitude of the relationships and display the peculiar reversal of relationships among superintendents, on the one hand, and teachers and board members, on the other. In virtually every large school system the superintendent rated himself higher than his associates rated him, and in virtually every small system he rated himself lower than the teachers and board members rated him.

The findings came as a surprise to Hunter since he had not expected any size differences in elevations of the LBDQ scores. His hypotheses dealt with the agreememt in LBDQ scores between the superintendent and his associates and not in the elevation of the scores. Specifically, he had predicted that agreement would be greater (or score discrepancies would be less) in small systems

Figure 1. Hunter's Data: LBDQ Scores from Superintendents, Teachers, and School Board members in large and small school systems.[a]

(Means of System Means)



**Significant beyond .01 level by t test, with 14 df. computed by author.

[a] From Hunter, op. cit., Tables VIII and X.

Table 1

Hunter's Data: Correlations between LBDQ Scores and Two Categories of School System Size, from Superintendents, Teachers, and School Board Members[a]

(Point Biserial Correlations)

	Initiating Structure	Consideration
Superintendent	.79	.68
Teacher	-.77	-.86
Board	-.66	-.47

---

[a] Correlations computed from data furnished by Hunter. Unequal variances in some instances prohibit the use of these coefficients for anything other than rough indications of the magnitude of relationships. All size differences except that for school board members on the Consideration dimension were statistically significant beyond the .01 level by appropriate forms of the t test. The exception did not reach significance at the .05 level.

where greater interaction occurred between the superintendent and teachers or board members than in the large school systems. He documented the fact, incidentally, of the difference in interaction rates between the two size-classes of school system. But his attempt to demonstrate the greater agreement in small school systems was confounded by the differences in elevations of the scores.

How does one account for the findings of Hunter? The LBDQ is supposed to be a measure of the performance of leaders. But the findings cannot be explained in terms of differences in leadership performance, since it is exactly the contradiction in this evidence which is to be explained. Perhaps small system superintendents are modest; maybe teachers in large systems dislike bureaucrats. Speculations of this sort require the assumption that LBDQ scores reflect attributes and circumstances of those who furnish descriptions of the superintendent as much as they reflect the superintendent's actual behavior. This is a reasonable assumption. But in order to explain the pattern of Hunter's findings, the attributes and circumstances of the raters--the superintendents, board members, and teachers--would have to vary systematically with school system size, and they would have to operate differentially among superintendents, on the one hand, and teachers and school board members, on the other. The possibility that the LBDQ scores unwittingly reflected such regularities among educational personnel and the organizational settings in which they were located gave impetus to the further exploration of Hunter's findings.

Another question must be raised and answered before it is worth giving serious thought to the nature of the forces acting upon school personnel. Hunter's findings were entirely unanticipated. Were they also accidental? The relationships of LBDQ scores with school system size may have been fortuitous--peculiar to the time and place in which the data were obtained. Were the relationships transient ones or are they generalizable to other school systems at other times? The LBDQ had been employed by other investigators to measure the behavior of superintendents and principals, and none had reported differences associated with school system size. Possibly the investigators had not looked for such differences. As marked and regular as Hunter's data showed the differences to be, it was remarkable that no one else had observed them if they were, indeed, general phenomena. Clearly, it was necessary to settle the question as to whether or not Hunter's findings could be replicated.

The Plan of Part I

Part I of this report describes the author's attempts to affirm the generality of Hunter's findings with respect to the differences in elevations of LBDQ scores associated with school system size. Prior to the full-scale replication of Hunter's study, supported by the U.S. Office of Education's Cooperative Research Program, the author inspected the data collected by other investigators. The remainder of the present chapter presents the results of these analyses as well as a re-analysis of Hunter's data which tested for a possible artifact attributable to questionnaire wording.

Chapter 2 describes the methodological procedures employed in the replication study. Details of Hunter's procedures are also presented there in order to facilitate comparisons with the replication study.

Chapter 3 reports the empirical data from the replication study, examined from several standpoints relevant to disparities in procedure between the two investigations. The final chapter in Part I briefly summarizes the significance of the results as they bear upon Hunter's original findings.

Attempted Confirmation: Data from Halpin

Shortly after Hunter's dissertation was completed, the author communicated the findings to Professor Andrew Halpin, who earlier had obtained descriptions of superintendents in 50 Ohio school districts by means of the LBDQ.[3] Of particular interest was whether or not Halpin had observed the same pattern of differences in LBDQ scores which Hunter discovered to be associated with school system size. Professor Halpin courteously furnished information regarding the pupil enrollment of each of the 50 school districts so that, in conjunction with the system-by-system LBDQ scores he had published, appropriate analyses could be undertaken. In the course of this correspondence, Professor Halpin also submitted his analysis of another set of LBDQ data which had just been collected in Illinois high schools.

---

[3] Andrew W. Halpin, Leadership Behavior of School Superintendents, School Community Development Study Monograph No. 4, Columbus, Ohio: Ohio State University Press, 1957.

Ohio superintendents. Like Hunter, Halpin had asked for descriptions of superintendents' leader behavior from school board members and from superintendents themselves. Where Hunter had collected descriptions from full-time classroom teachers, Halpin's third group of raters was what he called "staff". These were seven subordinates nominated by each superintendent as persons who worked closely enough with him to describe his behavior. They could be administrative assistants, non-certificated personnel, classroom teachers, or any combination of these. Hence, Halpin's "staff" respondents did not correspond precisely to Hunter's "teacher" respondents.

In another respect, the two studies were not exactly comparable. Hunter had changed the wording of the LBDQ items, which otherwise were identical with the items Halpin had used, so that they referred consistently and explicitly to the superintendent's behavior toward teachers. (Hunter's items are shown in Table 4.) In Halpin's version, the items referred either explicitly or implicitly to the superintendent's behavior vis-a-vis his immediate work staff. For example, Hunter modified the item, "He finds time to listen to staff members," to read, ". . . finds time to listen to teachers." While this modification entailed only a small number of word changes in the form, there is reason to believe that it introduced a markedly different context in which the ratings were to be made, an issue discussed later in this chapter and at length in Chapter 9.

Unfortunately, the Ohio school systems did not cover fully the same size range as Hunter's had. The systems in Halpin's study were predominantly rural and small in pupil enrollment, while Hunter had deliberately chosen metropolitan area systems of comparatively large size. Only one of the Ohio systems fell strictly within the limits of Hunter's "large" size category, and only five were even approximately equivalent. Fourteen of the Ohio systems corresponded to Hunter's "small" classification.

With these reservations regarding the comparability of the data, the analyses displayed in Table 2 do not lend support to the pattern of findings discovered by Hunter. Although four of the six differences shown in the table are in the proper direction to confirm Hunter's pattern, none are large enough to be regarded as anything but chance variations.[4]

---

[4] It should be noted that the six differences do not represent six independent tests. Initiating Structure and Consideration scores, both derived from the same subjects on the same questionnaire form (the fifteen items for each dimension were intermixed), are not methodologically independent. Measurement errors which vitiate the validity of the scores are likely to

Table 2

Halpin's Data: Mean LBDQ Scores from Superintendents, Board Members, and Staff in Large and Small School Systems

	Initiating Structure			Consideration		
System Size	Supt.	Board	Staff	Supt.	Board	Staff
Large	43.4	42.8	45.2	41.4	43.0	41.0
Small	43.1	46.7	44.0	36.7	44.9	39.1
Difference*	.3	-3.9	1.2	4.7	-1.9	1.9

---

*No differences reach statistical significance at the .05 level.

Illinois high school principals. Mentioned earlier was another set of data whose analysis had been furnished by Professor Halpin. These were LBDQ scores for principals in 40 Illinois high schools collected by Evenson for his dissertation research.[5] Descriptions were provided by the principals themselves, by their teachers, and by their superintendents. Information regarding size referred to high school size, not to school system size. Although Evenson's study had little in common with Hunter's other than its focus upon the leader behavior of school administrators, the data were checked for their systematic variation with size.

Table 3 displays the mean Initiating Structure and Consideration scores respecting the principal's behavior in the ten largest and the ten smallest high schools of the forty Evenson studied. It is obvious that no consistent size differences are revealed by the analyses.

The two attempts to find support for Hunter's pattern of findings in other studies were failures. Admittedly, neither of the two investigations were closely comparable to the Hunter

---

be reflected simultaneously in all 30 items of the questionnaire, rendering a correlation between the dimensions. Indeed, other studies have shown a modest correlation to be typical between the LBDQ dimensions. These same observations, of course, hold for Hunter's initial findings and apply to other analyses throughout this report.

[5] Warren Evenson, The Leadership Behavior of High School Principals: Perceptions and Expectations of Superintendents, Principals and Staff Members, Unpublished doctoral dissertation, University of Chicago, 1958.

Table 3

Evenson's Data: Mean LBDQ Scores Attributed to Principals by Superintendents, Principals, and Teachers in Large and Small High Schools

	Initiating Structure			Consideration		
System Size	Supt.	Principal	Teacher	Supt.	Principal	Teacher
Large	43.2	42.2	40.9	45.9	46.7	42.8
Small	43.1	40.3	41.0	49.8	46.9	45.3
Difference	.1	1.9	-.1	-3.9	-.2	-2.5

study, but their inspection indicated that Hunter's findings were not broadly generalizable to diverse educational settings and personnel.

Re-analysis of the Hunter Data

Early in 1960 the author conceived of a possible explanation for Hunter's observation of differences in teachers' descriptions of the superintendent in large and small school systems. Hunter had modified the original LBDQ items for his study by changing or adding words to indicate forms of leader behavior directed to teachers rather than to staff. Examination of these revised items suggested that some of them asked for ratings of behavior which was not within the realm of the superintendent's responsibility, particularly the responsibility of large-system superintendents. One of the items, for example, asked for a report of the frequency with which the superintendent "assigns teachers to particular tasks." In a small school system, with few levels of authority interposed between teacher and superintendent, the superintendent may in fact be expected to make such personnel assignments. In a large system, however, this responsibility normally rests in the hands of other administrators, usually building principals.

Since a number of the revised items were of this character, it appeared that the differences in the elevation of teachers' scores between large and small systems could be an artifact of the questionnaire modification. That some of the items did not seem to carry this "bias" against the large system superintendents offered a means of testing the explanation in Hunter's own data. It should be noted, however, that this explanation could account only for teachers' and school board members' differences in descriptions of the superintendent, not the differences in the opposite direction observed in superintendent self-descriptions. The item "bias" should have led large system superintendents to score themselves lower than small system superintendents, a result which patently did not occur.

The proposed explanation seemed to hold sufficient promise to warrant a test. Items of the LBDQ which were "biased" against the large-system superintendent should reveal the size differences observed by Hunter; those which were "unbiased" should show no differences associated with school system size. Accordingly, analyses were conducted on Hunter's responses from teachers with the intent, if the data supported the explanation, of extending the analysis to board member responses.

Preparation of a scoring key. Two graduate students and the author independently classified each of the 30 items in the Hunter revision of the LBDQ. The rationale for the classification was the normal locus of responsibility for the leader behavior described by the item--as resting with the superintendent or as resting with administrative personnel other than the superintendent (e.g., the principal) in typical public school systems. Where the locus of responsibility was judged to rest with the superintendent in small but not in large school systems, the item was classified as "biased" against the large-system superintendent. Where the locus of responsibility was judged equally likely to rest with the superintendent in small and large systems, the item was classified as "unbiased." No items appeared to be "biased" against superintendents of small districts, so this classification was not included.

The independent classifications by the three researchers showed unanimity on approximately one-half of the 30 items. Their discussion of disagreements revealed another rationale for distinguishing among the items. Some of the items referred to kinds of behavior which might ideally be displayed by a superintendent to teachers but which, because of the large number of teachers and the remoteness of the superintendent from teachers in large school systems, simply could not be displayed toward all or even many teachers. The following item from the Consideration dimension is a case in point:

"Gets teacher approval on important matters before going ahead."

Such behavior is more feasible for the superintendent of a school system with 30 teachers than for the superintendent of one with 300 teachers. Another item, apparently of the same kind, nevertheless was judged to be unbiased:

"Makes teachers feel at ease when talking with them."

In this case, the item refers to the behavior displayed by the superintendent in the event that a teacher talks with him and should be equally attainable by superintendents regardless of school size.

Discussion of disagreements permitted all but five of the 30 items to be classified. An item in the Initiating Structure dimension illustrates the dilemma of an item which could not unambiguously be regarded as biased or unbiased:

"Maintains definite standards of performance for teachers." On the one hand, the principal or supervisor in the large school system normally "maintains" and checks on the performance of teachers, but, on the other hand, the superintendent could be regarded as establishing and maintaining the criteria or standards by which teacher performance is judged.

It should be noted that none of the researchers had seen Hunter's data, other than total scores computed over all items, prior to their classification of the items by content.

About two-thirds of the 25 classifiable items were judged to be biased against the large district superintendent, and the biased items were equally distributed between the Initiating Structure and Consideration dimensions of the LBDQ, as shown below.

	Number of Items		
	Initiating Structure	Consideration	Total
Biased	8	8	16
Unbiased	4	5	9
Not Classified	3	2	5
Total	15	15	30

Table 4 displays the 30 items in the Hunter revision of the LBDQ, indicating the classification of each item as established through the foregoing procedures.

Results of the re-analysis. Raw data for the 240 teachers, furnished by Dr. Hunter, were transferred to Hollerith cards for processing. Regular scoring procedures were followed. Numerical values from 0 to 4 were assigned to response alternatives on each item to indicate low to high values on the Initiating Structure of Consideration dimension. Since there were unequal numbers or biased and unbiased items, "total" scores were computed over each set and divided by the number of items, yielding mean scores per item. Table 5 shows the results of the computations.

The entries of interest in Table 5 are the four difference figures between the large and small school systems. If Hunter's modification of the LBDQ could account for the size differences, the computations would show that teachers in the small systems rated their superintendents higher than those in large systems only on the 16 biased items and that little or no size differences would appear on the 9 unbiased items. It is apparent from the table that this did not occur. Size differences are as great in the unbiased as in the biased items. Thus, the data failed to support the proposed explanation, and no attempt was made to extend the analysis to responses of school board members.

To discover if any items were biased against the large systems, quite apart from those chosen on a priori grounds, the

Table 4

Items in Hunter's Revision of the LBDQ, Classified as Biased or Unbiased

Class*	Item
B	Does personal favors for teachers. (C)**
U	Makes his attitudes clear to the teachers. (IS)***
B	Does little things to make it pleasant to be a teacher. (C)
B	Tries out his new ideas with teachers. (IS)
NC	Is easy for teachers to understand. (C)
B	Rules teachers with an iron hand. (IS)
U	Finds time to listen to teachers. (C)
B	Criticizes poor work by teachers. (IS)
NC	Speaks to teachers in a manner not to be questioned. (IS)
B	Keeps to himself where teachers are concerned. (C)
B	Looks out for the personal welfare of individual teachers. (C)
B	Assigns teachers to particular tasks. (IS)
NC	Works without a plan. (IS)
NC	Maintains definite standards of performance for teachers. (IS)
U	Refuses to explain his actions to teachers. (C)
B	Acts without consulting his teachers. (C)
B	Is slow to accept new ideas from teachers. (C)
B	Emphasizes the meeting of deadlines for teachers. (IS)
NC	Treats all teachers as his equal. (C)
U	Encourages the use of uniform procedures. (IS)
U	Is willing to make changes concerning teachers. (C)
U	Makes sure that all teachers understand his part in the organization. (IS)
U	Is friendly and approachable for teachers. (C)
U	Asks that teachers follow standard rules and regulations. (IS)
U	Makes teachers feel at ease when talking with him. (C)
B	Lets teachers know what is expected of them. (IS)
B	Puts suggestions made by teachers into operation. (C)
B	Sees to it that teachers are working to capacity. (IS)
B	Gets teacher approval on important matters before proceeding. (C
B	Sees to it that the work of teachers is coordinated. (IS)

---

*B is Biased, U is unbiased, NC is not classified.

**Consideration dimension

***Initiating Structure dimension

Table 5

Teacher Scores on Biased and Unbiased LBDQ Items in Large and Small School Systems

(Per-item Means)

	Initiating Structure		Consideration	
System Size	**Biased**	**Unbiased**	**Biased**	**Unbiased**
Large	1.02	1.91	1.33	2.02
Small	1.48	2.35	1.77	2.52
Difference	-.46	-.44	-.44	-.50

teacher data were subjected to an item-by-item comparison of response distributions between large and small systems. The advantage of the small over the large districts, however, was remarkably uniform across all of the items, and this line of investigation was abandoned.

It is puzzling why such a reasonable hypothesis failed to be supported. Two reasons may be advanced. Teachers' descriptions of the superintendent's behavior may be governed less by the facts of the behavior than by their global, undifferentiated feelings about him. This possibility is discussed at length in Part II of the report. It is also possible that the role of the superintendent in the class of districts we have called "small" is not markedly different from that in the "large" systems, at least, not with respect to the kind of leader behaviors represented in the LBDQ. Chapter 9 raises the question of whether or not the LBDQ, originally designed to assess leadership in small, face-to-face work groups, is applicable at all to leadership in the formal organization of the school.

Plans for a Replication

Explorations by the writer between 1958 and 1960 had not resolved the questions which Hunter's unexpected findings had raised. Differences in teachers' descriptions of their superintendents did not appear to be artifacts of the questionnaire revision. On the other hand, there was no evidence to suggest that the same findings would hold in other settings with other size classes of school systems or with other kinds of respondents. The author's attempts to confirm the findings in other data had been unsuccessful.

At this juncture, the author was awarded a contract by the U.S. Office of Education under its Cooperative Research Program to replicate Hunter's investigation and, if the findings were confirmed, to attempt to account for them. The replication was to concentrate on teachers' reports of their superintendents and superintendents' self-descriptions, ignoring school board members altogether. This decision was taken on basis of the fact that differences in the board data in Hunter's study had been less impressive in magnitude than had those for the other two classes of school personnel. The study was designed to repeat Hunter's as closely as possible, introducing novelty only with respect to certain methodological precautions which Hunter had not observed and with respect to obtaining information which could serve to explain the findings in the event that Hunter's results were confirmed. The remainder of Part I of this report describes the replication of the 1958 investigation.

Chapter 2

Methods of the Replication Study

The research procedures necessary to duplicate Hunter's 1958 study[1] were not elaborate. He had selected 16 school systems in the general vicinity of St. Louis, Missouri, so as to constitute two size classes and had elicited Leader Behavior Description Questionnaire responses, principally by mail, from three categories of school personnel. Only a few controls were exercised in the selection of school systems to eliminate bases of variation other than size. His unanticipated findings were essentially zero-order correlations between the variables of size and the LBDQ scores.

The present study attempted to duplicate Hunter's research in its main respects, although one category of respondents--school board members--was not included. School systems in the same geographical area were chosen for study, including eight which Hunter had used three years previously. Only those procedural changes were introduced which provided additional information or which gave a stronger methodological base than in the 1958 investigation.

The first section of this chapter compares the 1958 and the 1961 studies with respect to four aspects of method and serves to describe the principal procedures of the replication study. The aspects of method are 1) the population of school systems used in the studies and certain differences in the size classes employed 2) procedures for selecting respondents from the school systems 3) methods of collecting data from respondents, and 4) the forms of the LBDQ. A second section of the chapter rounds out the procedural details of the replication study. Other matters of methodology not strictly relevant to the replication of Hunter's research are reserved for discussion in Chapter 6 of Part II.

---

[1] Ovid N. Hunter, Relationship between School Size and Discrepancy in Perception of the Superintendent's Behavior, Unpublished doctoral dissertation, Washington University, 1959. Our summary of his procedures in the present chapter is based upon his Chapter 2.

Methods in the 1958 and 1961 Studies

The site of the two studies was the metropolitan area of St. Louis, Missouri, covering highly urbanized suburbs of St. Louis as well as rural regions in the adjacent counties but excluding the City of St. Louis and its massive school system. The St. Louis Standard Metropolitan Area, designated by the U.S. Bureau of Census, encompasses St. Louis and Jefferson Counties in Missouri and Madison and St. Clair Counties lying across the Mississippi River in Illinois. School systems in these counties and within 35 to 40 miles of St. Louis were candidates for the investigations.

School systems and their size classes. While school systems were selected so as to vary in size, they were all required to meet the following criteria: 1) publically supported, 2) accredited by the North Central Association, 3) unit districts operating a 12-year educational program, and 4) employing superintendents who had held their positions for at least one full year. These represented the only explicit controls imposed upon the selection.[2] From among the eligible systems 16 were chosen for the 1958 study and 20 for the 1961 study.

	1958 Study	1961 Study
Number of "large" systems	8	10
Number of "small" systems	8	10
Total	16	20

The school systems differed with respect to student enrollment and, of course, with respect to all of the attributes associated with size, such as the number of faculty members, the number of building units, the size and complexity of the administrative operation, the financial base of support, and, to an extent, the urban as opposed to the small-town or rural character of the population served by the systems. The operational designations of "large" and "small" school systems is shown in Table 6 by the student enrollment figures for each of the 16 and 20 systems in the 1958 and 1961 studies respectively. "Large" systems means those with enrollments roughly between 5,000 and 10,000 students, although one of the systems in the 1961 study had nearly 20,000 students. "Small" systems means those enrolling somewhere between 500 and 3,000 students.

---

[2] To fulfill the requirements of a companion study, school systems in 1961 were also chosen to represent a range on a measure of "degree of bureaucratization." For details of the measure, see Gerald H. Moeller, The Relationship between Bureaucracy in School System Organization and Teachers' Sense of Power, Unpublished doctoral dissertation, Washington University, 1962. This consideration, however, did not constrain the selection of systems in filling the size categories; it was exercised only to choose between alternative systems when they already qualified for inclusion. The rank-order correlation between school system size and degree of bureaucratization for 20 systems was .43.

Inspection of Table 6 indicates that the large size class of systems in the 1961 study had a greater range than in the 1958 study, although its median size (7,900) was somewhat lower than in 1958 (9,200). The ranges and medians among the small systems in the two studies were more closely comparable. Despite the global similarities in size classes for the two studies, it is possible to select 11 systems from the 20 in the 1961 study to constitute size categories virtually identical to those of the 1958 study. This analysis is presented in the following chapter.

Of the 20 school systems in the 1961 replication study, 10 had been the sites of investigation in 1958, and in 8 of these (5 large and 3 small) the same persons still occupied the superintendent's position. Comparison of LBDQ scores in these 8 systems furnishes the closest replication possible.

Respondents and their selection. Both the 1958 and 1961 studies obtained questionnaire responses from superintendents and classroom teachers, but as noted earlier school board members were not included in the replication study. In the 1961

Table 6

Comparison of School System Sizes in Large and Small Size Classes of Hunter's Study and the Replication Study

(Number of Pupils)

	Hunter Study	Replication Study
Large		
	10,200	19,200
	10,100	11,500
	10,100	10,400
	10,050	10,300
	8,500	8,400
	7,800	7,400
	7,700	7,100
	6,500	6,700
		6,500
		5,500
Median	9,200	7,900
Small		
		2,800
	2,700	2,300
	2,600	1,800
	2,400	1,700
	1,600	1,500
	1,500	1,400
	1,500	1,000
	1,400	1,000
	750	900
		800
Median	1,550	1,450

replication, however, another category of school personnel was included--administrative staff members--principally as a means of establishing comparability with Halpin's investigation of Ohio school systems.[3] As reported in the preceding chapter, an analysis of Halpin's data for "staff" failed to confirm the relationship with school system size which Hunter had found for classroom teachers. But the two categories of personnel were not identical. There was a possibility that the size difference in LBDQ scores held true among classroom teachers but not among staff members. This possibility was checked in the replication study. Following Halpin's procedures, the 1961 investigator asked superintendents during interviews to name the three to seven certificated personnel (excluding full-time classroom teachers) with whom they worked most closely on a day-to-day basis.[4] The twenty superintendents named a total of 109 persons.

The selection of teacher respondents in the 1958 and 1961 studies entailed several points at which the procedures were not precisely alike. In both studies, a random sample was drawn from the roster of full-time classroom teachers in each school system, stratified by teaching level. The 1958 study employed three strata--elementary, junior high, and senior high school levels--and chose 6 teachers from each level. The 1961 study used only two strata--elementary and secondary--and drew 20 teachers from each. These figures are summarized below.

---

[3] Andrew W. Halpin, Leadership Behavior of School Superintendents, School Community Development Study Monograph No. 4, Columbus, Ohio: Ohio State University Press, 1957.

[4] The procedure did not correspond exactly to Halpin's. His superintendents could name any subordinate, including classroom teachers and non-certificated personnel.

1958 Study		1961 Study	
Elementary	6	Elementary	20
Junior High	6	Secondary	20
Senior High	6		
Total	18		40

(In some of the smallest systems of the 1961 study, the entire faculty consisted of only 35 to 45 teachers; in these cases, all teachers were included in the sample.)

The investigator departed from the stratification procedures of the 1958 study for several reasons. Hunter's plan of sampling junior high school teachers at the same rate as elementary and secondary teachers grossly over-represented them in the sample, and the general utility of the school system data was correspondingly limited. Sampling elementary and secondary teachers at the same rate, however, yields an approximately representative sample without resort to statistical corrections. Another reason was the difficulty of establishing a uniform definition of "junior high school" across all districts, some of which had no units by that name. The most compelling reason, however, was the investigator's demonstration, in a re-analysis of the Hunter data, that the findings with respect to size differences in LBDQ scores held for all classes of teachers, regardless of teaching level. Hence, differences in methods of stratification would have no bearing upon the results. These data are shown in Table I of the Appendix.

Another difference between the two studies lay in their measures for preserving the integrity of the randomization after samples were drawn and questionnaires were sent. Hunter regarded returns from a school system as "complete" when he had received the superintendent's response, returns from four of the six board members, and returns from five of the six teachers at each teaching level. Questionnaires in excess of these numbers were discarded as they arrived. Further, an indeterminant amount of replacement occurred in Hunter's teacher sample. He reported without elaboration, "When a blank or incomplete return was received with identification of the school system from which it came, another teacher was immediately contacted within the same school."[5] In some instances, entire school systems were dropped from his study when teacher or board member returns were insufficient. Several weeks after the initial mailing, Hunter sent reminder letters to teachers in the laggard systems, and after a month, a second questionnaire form was sent to all teachers in the groups which were still short of his cut-off point, this time to their home addresses. Hunter reported little information regarding the returns other than to indicate that his over-all return rate was 80 per cent, or 401 of 500 questionnaires distributed, "not including those sent later."[6]

---

[5] Hunter, op. cit., pp. 33-34.

The replication study, on the other hand, attempted to achieve as high a proportion of responses from the initial sample as possible, without recourse to replacement. Any self-selection bias of responding teachers would not be compounded by the investigator's procedures, as it could have been in Hunter's investigation. His data obviously were weighted in favor of teachers whose replies were prompt and thorough. While it is not clear how this weighting could produce the differences Hunter observed in LBDQ scores between large and small school systems, the replication study was designed to furnish a check on the possibility. A running account was made of the dates on which questionnaires were returned by teachers so that the prompt replies could be separated from the remainder. Analyses were conducted on these "first return" data in an effort to reconstruct Hunter's weighted sample.

To achieve a high rate of return of mailed questionnaires in the replication study, the investigator planned a series of follow-up procedures, including personal interviews with a sample of teachers who failed to reply after direct-mail appeals had been exhausted. This meant, however, abandoning the strict anonymity of response which Hunter could assure his teachers in his 1958 study. Since Hunter had no record of which teachers had and had not replied, except by school system and teaching level categories, he was obliged to direct his follow-up efforts to all teachers in the laggard categories. In the 1961 replication, letters of explanation which accompanied the questionnaires to teachers (and administrative staff members) called attention to a code number appearing in a corner of the form.

---

[6] Ibid., Table VI.

> This number allows us to determine which teachers have not returned questionnaires after a week or so in order that we may ask them again to do so. This is the sole use we will make of the code number. Once your questionnaire reaches us, your address card will be drawn out of our file and destroyed. In this way you may be assured of complete anonymity.

The letter also pointed out, in the customary manner, the purely research purpose of the investigation and the confidence in which all information from individual teachers would be held.

Data-taking methods. The central interest of the replication study, of course, was the response of teachers and superintendents on Hunter's version of the Leader Behavior Description Questionnaire. Hunter had used a two-page printed form containing, on one page, three items of identifying information and two questions regarding the nature and extent of teachers' (or board members') contacts with their superintendents. The 30-item LBDQ rating form filled the other page. The forms were identical for all three categories or respondent, although differential instructions were incorporated. Hunter distributed the forms personally to superintendents, through the superintendents to school board members, and through school offices to teachers; all were returned by mail.

In the replication study, several data-collection procedures were used. Superintendents received a simple, mimeographed LBDQ form accompanying the letter which notified them of the contemplated study. The investigator asked for an appointment to discuss the study in greater detail, and he requested that they complete the LBDQ form before the interview so that he could explain the purpose of the research fully and candidly. Superintendent responses were picked up from them during the interviews, except in four cases where they were not ready. These four questionnaires were returned later by mail, one of them so incomplete that LBDQ scores could not be computed.

Most of the interview time was devoted to an explanation of the research and its mechanics. Once the superintendent's permission was given (and no superintendent refused), additional information was obtained from him regarding 1) the amount of time he spent in contact with classroom teachers in his district, 2) persons on his administrative staff with whom he interacted most frequently, and 3) the names and home addresses of teaching personnel in the district. The latter information was furnished in the form of a school directory or roster.

LBDQ responses were obtained from administrative staff personnel--the individuals nominated by the superintendent--entirely by mail, using the same mimeographed form employed with superintendents. A covering letter explained the study, indicated that the superintendent had given his approval, and assured them anonymity in spite of the fact that an identifying number appeared on the form. As in the case of the teacher questionnaires, they were told that the number was to be used only for the purpose of following up respondents who had not returned their forms. Questionnaires were mailed to home addresses of respondents, accompanied by self-addressed, stamped envelopes for their direct return.

The same basic procedure was used to obtain responses from teachers as from administrative staff members, although the printed questionnaire form was more elaborate. The questionnaire contained three major parts, covering four pages, and thereby was longer than the one used by Hunter.

Part A, covering the first two pages, provided for 1) face-sheet information regarding age, sex, tenure in the system, teaching level, and the like, 2) an indication of the teacher's positions of responsibility other than classroom teaching, 3) a measure of friendship relations between the teacher and administrators and board members, 4) a measure of the frequency and nature of the teacher's contacts with the superintendent, and 5) a question relating to the existence of organized groups of teachers in the school system capable of influencing school system policies.

Part B contained the LBDQ items and, at the end of this page, a seven-point rating scale for evaluating the superintendent's leadership. As noted below, half of the teachers received a "personal" form of the LBDQ, and the other half received the "regular" form.

Part C, on the fourth page, contained a 30-item questionnaire incorporating five attitude scales. These scales were designed to measure the teacher's 1) sense of power with respect to school system policies, 2) affect toward the person of the superintendent (P Scale), 3) attitude toward upper-level administration (A Scale), 4) identification with the school system (S Scale), and 5) intrinsic satisfaction with the work of teaching (T Scale). The personal, affect-laden questions in these scales appeared after the LBDQ items to prevent their influencing, insofar as possible, the "set" by which teachers would respond to the LBDQ.

The Leader Behavior Description Questionnaire. The LBDQ form which served as the basis for Hunter's investigation was developed originally by the Personnel Research Board at Ohio State University.[7] In its early stages of development it consisted of 150 items designed to measure nine dimensions of leader behavior. Subsequent use of the questionnaire, particularly by Halpin, restricted attention to 30 items which provided measures of Initiating Structure and Consideration.[8] Halpin defined the two dimensions as follows:

---

[7] The developmental history of the LBDQ is presented in Ralph M. Stogdill and Alvin E. Coons, (Eds.) Leader Behavior: Its Description and Measurement, Research Monograph No. 88, Columbus, Ohio: Bureau of Business Research, Ohio State University, 1957.

[8] See Andrew W. Halpin and B. James Winer, "A Factorial Study of the Leader Behavior Descriptions," in Stogdill and Coons, (Eds.) Ibid., pp. 39-51; Andrew W. Halpin, Leadership Behavior . . . , op cit.

> Initiating Structure refers to the leader's behavior in delineating the relationship between himself and members of the work-group, and in endeavoring to establish well-defined patterns of organization, channels of communication, and methods of procedure. Consideration refers to behavior indicative of friendship, mutual trust, respect, and warmth in the relationship between the leader and the members of his staff.[9]

The 30 items in the form were divided evenly between the two dimensions and described behavior in which a school administrator might engage. The respondent's task was to check the frequency with which a given superintendent displayed each kind of behavior: "always," "often," "occasionally," "seldom," or "never." Weights of 0 to 4 were assigned to item alternatives, and a simple sum over items yielded scores on each dimension which could range from a low of 0 to a high of 60. Self-description forms varied only in the instructions. As reported in Chapter 1, Hunter changed the item wording by substituting teachers for staff. Hunter's item wording is shown in Table 4 of Chapter 1.

In the 1961 replication study, a randomly chosen half of the teachers in each school system received an LBDQ form which corresponded exactly to the one Hunter had used. The other half, however, received a "personal" form of the questionnaire, containing the same items but reworded to describe how the superintendent "behaves toward me" rather than how he "behaves toward teachers in general." Thus, the item "The superintendent of my school makes teachers as follows:

---

[9] Halpin, ibid., p. 4.

> Initiating Structure refers to the leader's behavior in delineating the relationship between himself and members of the work-group, and in endeavoring to establish well-defined patterns of organization, channels of communication, and methods of procedure. Consideration refers to behavior indicative of friendship, mutual trust, respect, and warmth in the relationship between the leader and the members of his staff.[9]

The 30 items in the form were divided evenly between the two dimensions and described behavior in which a school administrator might engage. The respondent's task was to check the frequency with which a given superintendent displayed each kind of behavior: "always," "often," "occasionally," "seldom," or "never." Weights of 0 to 4 were assigned to item alternatives, and a simple sum over items yielded scores on each dimension which could range from a low of 0 to a high of 60. Self-description forms varied only in the instructions. As reported in Chapter 1, Hunter changed the item wording by substituting teachers for staff. Hunter's item wording is shown in Table 4 of Chapter 1.

In the 1961 replication study, a randomly chosen half of the teachers in each school system received an LBDQ form which corresponded exactly to the one Hunter had used. The other half, however, received a "personal" form of the questionnaire, containing the same items but reworded to describe how the superintendent "behaves toward me" rather than how he "behaves toward teachers in general." Thus, the item "The superintendent of my school makes teachers feel at ease when talking with him," was converted to " . . . makes me feel at ease when talking with him." All but four of the 30 LBDQ items were changed in this respect, and the "personal" and "regular" forms were printed in otherwise identical questionnaires. (See Table II of the Appendix.) This procedure was limited to the forms for classroom teachers.

---

[9] Halpin, ibid., p. 4.

The alternative forms of the LBDQ were devised on the investigator's belief that the "regular" form used by Hunter was peculiarly susceptible to distortion by teachers' attitudes. Few teachers, especially in a large school system, would have any basis for knowing how the superintendent behaved toward "teachers in general." Confronted by questions of this sort, the teacher may attempt to answer from one of several frames of reference--how the superintendent relates to him personally, how he has treated other teachers in his circle of acquaintances, or how the superintendent is alleged to behave toward teachers in general. The ambiguity of the task encourages the teacher to project his own feelings and attitudes toward the superintendent or even toward other aspects of his work into his responses. The "personal" form of the LBDQ, however, does not leave the frame of reference in doubt. The teacher should be able to report how the superintendent behaves toward him as an individual, and his responses should be less subject to distortion. The investigator suspected that if the replication study confirmed Hunter's findings regarding school system size differences in teacher LBDQ scores on the "regular" forms, the differences would not appear on the "personal" forms.

Summary of method differences. Despite the over-all similarities, the replication study of 1961 departed from Hunter's study in several procedural details, some of which could conceivably affect the comparability of the LBDQ data. In some of these instances, however, checks were available in the 1961 study to assess the magnitude of effects. The four most significant departures which bear upon data comparability (other than the fact that no returns were obtained from school board members) are the following:

1. The size classes of the "large" and "small" systems were not exactly equivalent. The range in size of "large" systems was greater in the replication study, and the median size of both size categories was smaller than in Hunter's investigation. Analyses can be shown, however, for size classes reconstituted from among the 20 school systems which are virtually identical with Hunter's. Moreover, comparisons of responses are available in 8 school systems which were included in both studies and which were still governed by the same superintendent.

2. The sampling of teachers in the replication study was conducted without replacement. Again, it is possible to check the difference in LBDQ responses which would have occurred had the replication study relied only upon teachers who returned their questionnaires promptly and fully completed. The "first return" responses in the replication study approximate Hunter's replacement sample.

3. Code numbers on the questionnaires certified that no respondents in the replication study could be utterly anonymous, whereas in Hunter's study the teacher questionnaires were unidentified. Presumably, the cloak of anonymity would allow teachers to be more derogatory in their descriptions of superintendent behavior and scores would be generally lower in the replication study than in the 1958 investigation. This matter can be examined readily. In both studies, however, superintendent self-responses were equally non-anonymous, and no differences in their scores could be attributed to the factor.

4. Teacher questionnaires in the replication study measured more variables of a more personally probing sort than had Hunter's. Most of these probing questions were placed on the last page of the questionnaire form so that, if teachers answered the questions in the order of their appearance, the LBDQ responses would be unaffected.

Additional Details of Procedure: 1961 Study

The active phase of the replication study covered most of the academic year of 1960-61. In the fall of 1960, attitude scales were constructed by the research staff and pretested on 125 teachers in extension courses at Southern Illinois University and in a St. Louis area school system which did not qualify for the replication study proper. Initial contacts with the superintendents of the 20 school systems were made just before and just after the

Christmas holidays, at which time their nominations of staff members and their self-description LBDQ forms were obtained.

During the month of February, 1961, 109 questionnaires for members of the administrative staffs and 783 teacher questionnaires were mailed to home addresses, accompanied by explanatory letters and envelopes for their return by direct mail. The time of year in which respondents completed the forms was the same as in the Hunter study.

As questionnaires were returned to the investigator's office, the date of receipt was stamped on the form, and a running account was kept of the returns from each school system. When the number of returns tapered off appreciably, usually within two weeks, a letter and a duplicate copy of the questionnaire was sent to non-respondents. Plans for a second follow-up letter and/or personal contacts with non-respondents were abandoned when the response rate reached an unexpectedly high level: by the end of April, 1961, the rate of return for teachers was 88 per cent of the original sample and 96 per cent for the administrative staff members. While response rates varied substantially from one school system to another--from a low of 74 per cent to a high of 100 per cent--the rates were essentially the same for large and for small size classes of system, being 88 and 89 per cent, respectively.

When no more questionnaires arrived, the research staff transferred responses to tabulating sheets and a commercial tabulating service key-punched the data on Hollerith cards. Check-coding of a 10 per cent sample of questionnaires indicated that the transfer of data to tabulating sheets entailed a coding error of less than 1 per cent, insufficient to warrant re-coding of all questionnaires. A standard verification procedure was followed by the commercial tabulating service, eliminating virtually all errors in that phase of the operation.

At the end of May, 1961, feedback reports were mailed to each school system, based upon special tabulations of data which would be of interest in the system. The tabulations permitted comparisons of mean scores among various categories of personnel in the system and between the particular system and the nine others in the same size category--with regard to the LBDQ dimensions and the attitude measures. The other school systems, of course, were unidentified.

Unusable returns. The total sample of teachers was 783, from whom 692 replies of some sort were received. Thirty of these teacher replies, however, were ineffective insofar as the study was concerned. Three teachers claimed they had returned questionnaires, but the investigator's office had no record of their receipt. Another five respondents reported that they either were no longer at the school system or were not full-time classroom teachers. In 22 cases, questionnaires were returned blank (usually with an accompanying note indicating that the respondent preferred not to fill it out) or so incomplete that there was no point in preparing Hollerith cards for them. Thus, the effective N of the teacher sample was 662, distributed almost exactly evenly between large and small school sytem categories.

While 662 teacher responses were sufficiently complete to warrant their inclusion in the study, not all of them could be scored on one of the other (or both) of the LBDQ dimensions. A strict criterion of scorability was maintained by the research staff: all 15 questions for a dimension had to be answered and answered unambiguously before a score was computed. Otherwise, the respondent's score on the dimension was coded as Not Ascertained. On this basis, 169 questionnaires could not be scored on the Initiating Structure dimension and 163 could not be scored on Consideration; most of the unscorable LBDQ dimensions were in questionnaires of the same teachers.

The figures below summarize the statistics regarding the total usable replies from teachers and the scorable LBDQ responses, combining the "regular" and "personal" forms.

	Number	Per Cent of Sample
Total teacher sample	783	100%
Responses	692	88
Usable returns	662	85
Scorable LBDQ dimensions		
Initiating Structure	493	74
Consideration	497	75

There were no differences in the number of usable returns from teachers receiving "regular" or "personal" forms of the LBDQ--328 and 334, respectively--although somewhat lower proportions of the "regular" forms were scorable.

For the 109 administrative staff members initially nominated by superintendents, responses were received from 104. Four respondents omitted some items on one or the other LBDQ dimension--two on each dimension--so that tabulations were based upon an N of 102. The return statistics for staff members follow:

	Number	Per Cent of Total
Total administrative staff	109	100%
Responses	104	96
Scorable LBDQ dimensions		
Initiating Structure	102	94
Consideration	102	94

As mentioned earlier, one superintendent of the 20 from whom self-descriptions were requested returned the form so incomplete that no score on either dimension could be computed, not even approximately.

Attributes of non-respondents. While 85 per cent of the teachers provided usable questionnaire returns, 12 per cent did not respond at all and 3 per cent replied but furnished no usable data. Limited information regarding the non-respondents was available from the school rosters from which their names were taken. Comparisons shown in Table 7 between the 91 teachers who did not reply ("non-returns"), the 30 who replied but furnished no usable information ("unusable returns"), and the 662 who replied with sufficient data to tabulate ("usable returns") demonstrate several differences. The respondents upon whose data statistical tabulations could be made were more likely to be 1) from small rather than large school systems, 2) males rather than females, and 3) teaching at the secondary rather than the elementary level. Unknown, of course, is the extent to which respondents differed from non-respondents with regard to their descriptions of leadership behavior of superintendents, their attitudes, and other attributes relevant to the present investigation. Data are presented later which suggest that differences between respondents and non-respondents in LBDQ scores are unlikely to be substantial, but these data are not conclusive.[10] The small number of non-respondents, however, makes it impossible for their exclusion to affect appreciably the results of the replication study, even were they markedly different from the respondents in their attributes.

Summary

To replicate Hunter's 1958 investigation, the 1961 study obtained self-descriptions from superintendents and descriptions from teachers of the superintendents' leadership behavior in 20 school systems of the St. Louis metropolitan region. No data

---

[10] Teachers who replied immediately to the questionnaire mailings ("first returns") did not differ in the LBDQ scores attributed to superintendents than teachers who replied only after further urging ("subsequent returns"). See infra, Chapter 3.

Table 7

Sex and Teaching Level Distributions of Responding and Non-Responding Teachers

	Respondents (Usable Data)	Respondents (Unusable Data)	Non-Respondents
Sex			
Male	34%	23%	30%
Female	66	77	70
NA	*	-	-
Total	100%	100%	100%
(Number)	(662)	(30)	(91)
Teaching Level			
Elementary	46%	50%	53%
Junior High	19	13	14
Senior High	34	37	33
NA	1	-	-
Total	100%	100%	100%
(Number)	(662)	(30)	(91)

---

*Less than 1 per cent.

were taken from school board members, but LBDQ responses were obtained from members of the superintendents' administrative staff in order to provide overlap with Halpin's study of Ohio school superintendents. The same criteria were used for selecting schools as Hunter had applied, and with certain qualifications the size categories into which the school systems fell--in terms of pupil enrollment--were approximately the same in the two studies. Eight of the systems were the same as in the Hunter study, still with the same person as superintendent.

A minor departure in stratification procedures for sampling teachers and a more important one with regard to preserving the integrity of the random sample in the questionnaire returns occurred in the 1961 study. Where Hunter had sent questionnaires to additional teachers to replace those who did not reply, the 1961 study used vigorous follow-up methods to assure a high rate of return from the original sample. This entailed, however, breaking the strict anonymity of respondents which Hunter had observed. Replies were received from 88 per cent of the 783 teachers, 96 per cent of the 109 administrative staff members, and 19 of the 20 superintendents in the 1961 study.

Questionnaires to superintendents and administrative staff members contained only the LBDQ items, but those to teachers included two pages of face-sheet questions and a final page consisting of 30 items measuring five kinds of teacher attitude, as well as the LBDQ questions. Many of the additional questions in the teacher questionnaire were designed to account for differences in teacher descriptions of superintendents in the event that Hunter's findings were confirmed. For the same reason, two forms of the LBDQ were used--one the "regular" form identical to the one Hunter employed and the other a "personal" form in which items were reworded to refer to how the superintendent behaved toward the responding teacher himself rather than toward teachers in general These alternative forms were distributed to random halves of teachers in each school system in otherwise identical questionnaires.

Data were taken from respondents in February and March of 1961, the same time of year as Hunter had collected his. After data were transcribed, special feedback reports were prepared for each cooperating school system.

Chapter 3

Results of the Replication Study

The critical question for the replication study is whether or not the *pattern* of differences observed by Hunter in 1958 in superintendent self-descriptions and teacher descriptions of superintendent leader behavior would re-appear in a new study. Specifically, Hunter's study had shown significant *positive* correlations between superintendents' self-descriptions of Initiating Structure and Consideration and school system size and significant *negative* correlations between teacher descriptions and school system size.

This chapter assembles the relevant data respecting the pattern of LBDQ scores associated with size. Comparisons are presented first for all 20 school systems of the two size categories and then for 11 systems selected to correspond almost exactly to the size classes of Hunter's study. Responses furnished by superintendents and teachers in 8 school systems which were studied in both 1958 and 1961 are compared for their consistency over the three-year period. In addition, comparisons are made in the responses of teachers who replied promptly to the request for superintendent descriptions (and who, thereby, approximate Hunter's "biased" sample) and in teacher responses on the "personal" form of the LBDQ. LBDQ scores for each school system in the replication study are presented in Table III of the Appendix.

Comparisons of LBDQ Scores by System Size

Table 8 shows the point biserial correlations between school system size and Initiating Structure and Consideration scores among superintendents and teachers in the 20 school systems.[1] Only one of the four relationships reached the 5 per cent criterion of statistical significance. Contrary to Hunter's finding, teachers in large systems scored their superintendents significantly higher on Initiating Structure than did those in small systems.[2] But even this relationship, .45, was small in absolute magnitude compared with the realtionship of -.77 which Hunter had observed. More importantly, however, the pattern of relationships found by Hunter was not duplicated in the 1961 study. The only case in which the direction of the relationship is the same as Hunter had found is that of superintendent scores on the Consideration dimension, but it does not depart significantly from zero. The most striking feature of the mean scores (Table 9) is the minute order of the differences between large and small systems, whatever their direction.[3]

---

[1] These data are for the "regular" forms of the LBDQ only.

[2] Tests of statistical significance were computed not on the point biserial correlations but on the mean scores for the size categories (Table 9) where appropriate forms of the *t* test could be applied to take into account inhomogeneity of variance.

[3] The data for teachers in Table 9 are the means of mean school system scores. In preparing these data for presentation, the mean of teacher scores within each school system was computed, yielding a system score; then the means of the system scores were computed for the ten school systems in each size class. An alternative procedure would be to compute mean scores for all teachers in each

Table 8

Correlations between LBDQ Scores and Two Categories of School System Size from Superintendents and Teachers

(Point Biserial Correlations)

	Initiating Structure	Consideration
Superintendent	-.13	.17
Teacher	.45*	.22

---

*$p <$ .05, based upon a t test of means shown in Table 9.

Table 9

Mean LBDQ Scores from Superintendents and Teachers in Large and Small School Systems

	Initiating Structure		Consideration	
System Size	Superintendent	Teacher	Superintendent	Teacher
Large	36.4	38.6	45.0	41.3
Small	38.0	35.0	43.3	38.7
Difference	-1.6	3.6*	1.7	2.6

---

*p $< .05$

The mean scores for superintendents and teachers, when compared with the means Hunter obtained in his 1958 study, bear upon the question regarding the effects of anonymity in questionnaire responses. Teacher responses in 1961 were not entirely anonymous, and teachers might be expected to furnish more favorable descriptions of their superintendents (higher LBDQ scores) under this condition than under Hunter's condition of complete anonymity if, indeed, anonymity affects responses at all. In both studies, superintendent responses were equally non-anonymous and should show no differences in LBDQ elevations. Table 10 compiles the data necessary for these comparisons.

The differences between the 1958 and 1961 data shown in Table 10 are not consistent with an interpretation that LBDQ scores are affected by anonymity. Superintendent scores vary at least as much as teacher scores do, although the conditions of anonymity were the same for them. Since superintendent scores are based on

---

size class directly, without going through the intermediate step of finding system scores. Both Hunter and the present investigator computed means according to the more direct procedure as well as in terms of school system scores, but only the latter are displayed in this report.

Similarly, both Hunter and the present investigator computed tests of statistical significance on system scores and on the pooled scores of all teachers within the size classes, correcting for inhomogeneity of variance where necessary. Practically speaking, however, it made no difference in either Hunter's data nor in the data of the replication study how tests of statistical significance were run. Where size differences were significant under one test, they were significant under the other; where insignificant differences appeared under one, so were they insignificant under the other.

Table 10

Comparison of Mean LBDQ Scores in the 1958 and 1961 Studies, from Superintendents and Teachers in Large and Small School Systems

	Large		Small	
	Initiating Structure	Consideration	Initiating Structure	Consideration
1961 Supt.	36.4	45.0	38.0	43.3
1958 Supt.	47.6	50.5	38.2	42.9
Difference	-11.2	-5.5	-.2	.4
1961 Teachers	38.6	41.3	35.0	38.7
1958 Teachers	35.4	39.3	42.8	46.4
Difference	3.2	2.0	-7.8	-7.7

small numbers of cases compared with the teacher scores and, hence, could be expected to be more variable, this evidence is not as convincing as the examination of teacher scores alone. In the case of teachers, the mean scores were lower, not higher, under the 1961 condition of quasi-anonymity than under the condition of full anonymity. That is, teachers rated their superintendents less favorably when they had reason to suspect (despite our assurances to the contrary) that their responses would be identified. Further inspection of the scores demonstrated this tendency to be markedly stronger among teachers in the small than in the large school systems--among those whom one would expect to be the most threatened by the removal of the cloak of anonymity. (These data are not shown.) Thus, it is unreasonable to attribute differences in LBDQ scores to variations between the studies in conditions of anonymity.

Curiously, an important size difference appeared in administrative staff members' ratings of their superintendents. As Table 11 demonstrates, staff members in the large systems rated superintendents higher on both the Initiating Structure and the Consideration dimensions than those in small systems. Hunter, of course, had no data against which these can be compared, but it is noteworthy that they do not conform in direction with Hunter's findings for teachers. Also, the correlations of .43 and .44 are substantially lower in absolute magnitude than most of those observed by Hunter (Table 1).

Table 11

Correlations between LBDQ Scores and Two Categories of School System Size from Administrative Staff Members

	(Point Biserial Correlations)	
	Initiating Structure	Consideration
Administrative Staff	.44*	.43*

---

*$p < .05$, computed by an appropriate t test on LBDQ mean scores.

Differences in Adjusted Size Classes

The preceding chapter indicated that the size classes of school systems used in the replication study did not conform exactly to those of Hunter.[4] It is possible, however, to adjust both the ranges and medians of student populations among the ten large and ten small systems by eliminating selected school systems from the replication study. By discarding five large systems and four small systems, the ranges and medians can be made more closely similar between the two studies. Systems 01, 02, 05, 06, and 09 were removed from among the large size category and systems 11, 15, 16, and 18 were removed from the small size category, resulting in the following comparisons of student populations for the two studies:

	Range	Median	Number
Large, 1958	6,500-10,200	9,200	8
Large, 1961	7,100-10,400	8,400	5
Small, 1958	800-2,700	1,600	8
Small, 1961	1,000-2,300	1,600	6

Although some discrepancies still exist, they are slight.

LBDQ scores for superintendents and teachers, computed in the school systems which remain after adjusting the size classes, still fail to reveal the pattern of differences reported by Hunter. (See Table 12.) Only with respect to teacher scores on the superintendents' Initiating Structure is there a reasonably large dif-

---

[4] See pp.24-26 and Table 6, supra.

Table 12

Mean LBDQ Scores from Superintendents and Teachers in Adjusted Size Categories of Large and Small School Systems

System Size	No. Systems	Initiating Structure	Consideration
		Superintendent	
Large	5	36.0	44.2
Small	6	37.4	44.0
Difference		-1.4	.2
		Teacher	
Large	5	39.2	42.0
Small	6	34.9	41.4
Difference		4.3	.6

ference, similar to that found in the earlier analyses. Again, it is in favor of the large system superintendents rather than the small system superintendents in contrast to Hunter's finding. (Tests of statistical significance were not computed for these data.)

Same System Comparisons

As mentioned previously, some of the school systems in the 1961 replication study were the same as Hunter had studied three years previously, still with the same persons as their superintendents. Five school systems in the large size class and three in the small size group are available for the direct comparison, although the superintendent in one of the latter systems furnished deficient LBDQ data and his scores could not be computed. The circumstance of having the same school systems in the two studies provides an opportunity to examine the stability in LBDQ scores over the three-year period as well as to test, in the most closely comparable fashion possible, the generality of Hunter's findings. Insofar as Hunter's original study is concerned, these eight systems showed the same pattern of size differences as his full sixteen: superintendent self-descriptions were markedly higher and teacher descriptions markedly lower in the large systems than in the small. (See Tables 13 and 14.)

Superintendent self-descriptions. The striking size differences in superintendent self-descriptions found by Hunter virtually disappeared in the replication study among these identical systems. Where Hunter had found a mean Initiating Structure score of 47.0 and 39.5 in the five large and two small systems, respectively, the means on Initiating Structure in the replication study were 38.2 and 38.0, respectively. Hunter's data show a separation of 8.9 between the Consideration means of large and small systems; in the replication study this difference reduced to 3.8. (Table 13.)

It is noteworthy that the correlations between the superintendents' scores over the three-year period are positive but low. The rank difference correlation, computed for superintendents in all seven systems, is .27 on Initiating Structure and .38 on Consideration, where .79 is required for significance at the 5 per cent level. The range of superintendent scores remained about the same over the period. Thus, it is impossible to predict on better than a chance basis the standing of the superintendent in the replication study from his stnading in the Hunter study.

Teacher descriptions. Teacher descriptions of their superintendents in the five large and three small systems in 1961 also fail to reveal the size differences in Initiating Structure and Consideration scores which Hunter had found earlier. Where system scores in 1958 were 4.8 points higher on Initiating Structure and 8.8 points higher on Consideration in the small systems than in the large, in 1961 the scores differed at the most by 1.7 points (on Initiating Structure) and in a direction favoring the large systems instead of the small. (Table 14.) They are not statistically significant.

Table 13

LBDQ Scores from Same Superintendents in 1958 and 1961

	Initiating Structure		Consideration	
System Code	1958	1961	1958	1961
Large				
01	52	49	54	55
05	49	30	55	45
07	50	45	54	49
08	42	34	44	36
09	42	33	40	39
Mean	47.0	38.2	49.4	44.8
Small				
11	41	31	42	35
13	38	45	39	47
Mean	39.5	38.0	40.5	41.0
Differences in Means, Large-Small*	7.5	.2	8.9	3.8
Rank difference correlation, 1958-1961 (all 7 systems)*	.27		.38	

---

*Neither the differences in means in the 1961 data nor the correlation coefficients are significantly different from zero at the .05 level.

Table 14

Mean LBDQ Scores Attributed to Same Superintendents

by Teachers in 1958 and 1961

	Initiating Structure		Consideration	
System Code	1958	1961	1958	1961
Large				
01	37.1	38.3	38.4	41.1
05	32.3	40.1	37.3	44.2
07	41.1	43.4	39.2	41.1
08	37.7	40.8	43.5	43.6
09	34.5	36.5	36.9	36.8
Mean	36.5	39.8	39.1	41.4
Small				
11	39.7	36.0	48.0	41.4
12	41.5	35.2	46.5	38.2
13	42.6	43.0	49.3	43.8
Mean	41.3	38.1	47.9	41.1
Difference in Means, Large-Small*	-4.8	1.7	-8.8	.3
Rank difference correlation, 1958-1961 (all 8 systems)*	.38		.30	

---

*Neither the differences in means in the 1961 data nor the correlation coefficients are significantly different from zero at the .05 level.

Rank difference correlations over all eight systems between 1958 and 1961 showed a .38 relationship on Initiating Structure and a .30 relationship on Consideration. A .79 correlation is required for significance at the 5 per cent level. A slight constriction in the range of system scores is observed in the replication study, but probably not so great as to reduce the opportunity to obtain a significant correlation. The range of system scores for teachers in the eight systems had been 10.3 and 11.1 on Initiating Structure and Consideration respectively; these reduced to 8.2 and 7.3 respectively in the replication study.

It is fully apparent that Hunter's 1958 findings were not confirmed by the replication study at the point where the opportunity for confirmation was the greatest--in the school systems identical to those which Hunter had studied.

Moreover, the comparisons indicated a high degree of instability in both teacher and superintendent LBDQ responses over the three-year period. None of the rank difference correlations between the two years was statistically significant. It is possible, of course, for superintendents to modify their modes of leadership behavior in three years' time and thereby render the descriptions unstable. Additional computations, however, suggest that the instabilities are not simply results of changes in the superintendents' behavior. Rank order correlations between the 1958-1961 differences in teacher descriptions and the 1958-1961 differences in superintendent descriptions do not depart from zero except by chance.

(These computations are not shown but may be derived from Tables 13 and 14.) If anything, the correlations indicate that when superintendents report higher scores on leadership behavior, teachers are inclined to report lower scores for them, and vice versa. The rank difference correlations between the differences were -.36 for Initiating Structure and -.53 for Consideration, where a coefficient of .79 is required for statistical significance at the 5 per cent level.

Size Differences in the Early Returns

One of the procedural differences between the 1958 study and the 1961 replication lay in the selection of teacher responses from the school systems. As reported in Chapter 2, both studies used a random procedure to obtain their initial lists of teacher respondents to whom questionnaires were distributed,[5] but in the replication study no further selection was made by the investigator. On the other hand, Hunter chose for his analysis only the first fifteen complete questionnaires to arrive from a given school system, discarding any above that number. In those cases where the initial list did not yield fifteen complete returns, he circulated questionnaires to additional teachers. Thus, in some unknown degree Hunter destroyed the randomness of his samples; the returns used in his analyses over-represented teachers who were prompt in sending back complete questionnaires. This was not a crucial matter for the hypotheses which he wished to test in his study, but if the attributes of these prompt teachers somehow were associated with inclinations in describing the leader behavior of the superintendent, the over-representation might account for his findings regarding elevations in teacher LBDQ responses.

---

[5] In a few of the smallest systems of the replication study, the entire population of classroom teachers was included.

It is possible to approximate Hunter's biased sample in the present data by restricting the analysis of teacher LBDQ scores to questionnaires which were returned immediately, without further urging by the investigator. Records had been kept regarding the date of questionnaire returns, and those which had arrived within the 15-day period before follow-up letters were dispatched to the laggard teachers were chosen for this analysis. Of the 328 usable returns eventually received from teachers, 218 (or 66 per cent) arrived within the 15-day period. The percentage of early returns was somewhat higher in the small systems (70 per cent) than in the large (63 per cent). Not all of the early returns could be scored on the LBDQ dimensions, however, since some teachers either did not complete this part of the questionnaire or skipped or otherwise spoiled certain items within the LBDQ form. This occurred more frequently in responses of teachers from large than from small systems. There remained 172 questionnaires from early-responding teachers which furnished complete scores on the Initiating Structure dimension and 176 which were complete on the Consideration dimension. These returns should approximate the biased sample upon which Hunter based his analyses.

Computations of mean Initiating Structure and Consideration scores for early respondents in the large and in the small school systems revealed no differences such as Hunter had found. Table 15 shows these means, in this instance computed from the pooled scores of all teachers in the size classes rather than from system scores. It is the teachers in the large systems, not the small systems, who impute higher incidences of Initiating Structure and Consideration to their superintendents. As Table 15 also indicates, the LBDQ scores derived from the early respondents differ only microscopically from the scores of all respondents in the replication study. In other words, there is no evidence of an association between attributes of early responders and inclinations in describing superintendents.

Size Differences in the "Personal" Form of the LBDQ

Chapter 2 described the construction of a "personal" form of the LBDQ and its distribution to a random half of the teachers in each school system. Less ambiguous than the "regular" form in question wording (or, at least, in the frame of reference teachers could adopt in their replies), the "personal" form was originally included as a means for explaining the Hunter findings in the event they were confirmed by the replication study. While it is clear that the findings were not confirmed, these data nevertheless were examined for their associations with school system size.

Mean LBDQ scores on the "personal" form tended to be lower than those on the "regular" form, but, as Table 16 demonstrates,

Table 15

Mean LBDQ Scores from Teachers Who Replied Early Compared with All Teachers in Large and Small School Systems

System Size	Initiating Structure		Consideration	
	Early Returns	All Returns	Early Returns	All Returns
Large	38.5	38.7	41.7	41.3
Small	34.6	35.1	38.1	38.1
Difference	3.9	3.6	3.6	3.2

Table 16

Mean LBDQ Scores from Teachers in Large and Small School Systems, "Personal" Form of the LBDQ

System Size	Initiating Structure	Consideration
Large	32.2	34.7
Small	29.5	37.8
Difference*	2.7	-3.1

---

*No differences reach statistical significance at the .05 level.

the differences between teacher descriptions of superintendents in large and small school systems are inconsistent in direction and, in any event, not statistically significant. Only with respect to Consideration do teachers in the small systems rate their superintendents higher than those in large systems, but the statistical insignificance of the difference prevents it from constituting partial confirmation of Hunter's findings.

It is notable that the significantly higher Initiating Structure scores attributed to large system superintendents by their teachers on the "regular" form of the LBDQ (Table 9) no longer represents a statistically significant difference on the "personal" form responses. When teachers were asked to describe the superintendent's behavior toward the vague class of "teachers in general," large system superintendents received higher scores than small system superintendents' but when teachers were limited to descriptions of the superintendent's behavior toward them individually, differences associated with school system size did not appear.

Summary

No matter how the data of the replication study are inspected, no support is forthcoming for Hunter's unanticipated findings in 1958 that superintendents in large school systems rate themselves higher on two dimensions of leader behavior than their counterparts in small systems and, conversely, that teachers in large systems rate their superintendents lower in leader behavior than teachers in small systems.

1. No size differences appeared in superintendents' scores.

2. Teachers in large systems rated their superintendents higher on Initiating Structure than those in small systems, contrary to Hunter's observations; no statistically significant difference appeared with respect to teacher descriptions of the superintendents' Consideration behavior.

3. Administrative staff members in large school systems rated their superintendents significantly higher on both Initiating Structure and Consideration than those in small systems. Hunter's study furnished no data against which this finding could be compared.

4. When certain school systems in the replication study were discarded so that the remaining systems would conform more closely to Hunter's size categories, the LBDQ scores for superintendents and teachers did not manifest the size differences observed by Hunter.

5. Teacher and superintendent scores in precisely the same school systems Hunter had studied three years previously, still served by the same superintendents, failed to show the striking size differences which had appeared in these systems earlier. Rank difference correlations of scores between 1958 and 1961 did not differ significantly from chance, demonstrating an absence of stability in superintendent self-descriptions and teacher descriptions of the superintendent over the three-year period.

6. Restricting the analysis only to teachers who returned their questionnaire forms promptly and in complete form, thereby

approximating the conditions of Hunter's sample, the anticipated size differences did not appear. Scores on these early returns did not differ appreciably from scores on the total set of returns, suggesting that promptness was unrelated to descriptions of the superintendent.

While no rationale was advanced for the analysis, scores from the "personal" form of the LBDQ, circulated to a random half of the teacher respondents, were examined for indications of size differences. None were found.

Hence, only three statistically significant differences in LBDQ scores associated with school system size emerged from the various tests, two of which cannot be compared with any of Hunter's findings (administrative staff members in large systems rated their superintendents higher on both Initiating Structure and Consideration than staff members in small systems) and the other of which was contrary to Hunter's data. Teachers in large systems rated their superintendents significantly higher on Initiating Structure in the "regular" form of the LBDQ than did teachers in the small systems, but even this contrary difference did not stand up when teacher responses on the "personal" form of the LBDQ were considered. In short, the replication study unequivocally disconfirmed the findings from Hunter's 1958 investigation.

Chapter 4

Conclusions of the Replication Study

Data presented in the previous chapter demonstrate clearly that the unanticipated findings which emerged from the 1958 dissertation study by Hunter were disconfirmed. Why should marked differences appear in one study and then fail to appear in a second study conducted under highly comparable circumstances?

In Chapter 2, four points were noted in which the replication study departed methodologically from the 1958 investigation. These had to do with (1) non-equivalent size classes, (2) procedures for sampling teachers, (3) anonymity of replies, and (4) differences in questionnaire format. In Chapter 3, data were brought to bear upon three of these four points. Certain systems in the replication study were eliminated to render the size classes closely similar, but Hunter's differences in LBDQ scores did not appear. Analyses were conducted on only those questionnaires which were returned promptly, a restriction which approximated the conditions of Hunter's sample, but no size differences in LBDQ scores were observed among them. With respect to anonymity, teacher descriptions of the superintendent were not more favorable under the conditions of quasi-anonymity of the 1961 study than under conditions of full anonymity of the 1958 study. In fact, just the reverse occurred, and especially so in the small school systems where nonanonymity would be least likely to result in unfavorable superintendent ratings.

Regarding the questionnaire format, the LBDQ items were part of a longer questionnaire in the replication study in which rather penetrating questions were asked of teachers about their reactions to the circumstances of their work and their personal evaluations of the superintendent. For the most part, these questions followed the LBDQ items on an over-leaf. If respondents went through the questionnaire in the order in which the items appeared (which, of course, they need not have), their responses to the LBDQ items would be unaffected. Hunter's questionnaire did not contain any such questions. The investigator cannot conceive of any reason, though, why the inclusion of these additional questions could have the specific effect of lowering the LBDQ responses of teachers in the small systems and/or raising those of teachers in the large systems, thereby eliminating the size differences observed by Hunter.

The procedural differences between the 1958 study and the 1961 study are incapable of explaining the presence of findings in the first and the absence in the second.

A fundamental issue of research strategy arises at this point. How far should one go in taking as problemmatic the fact that one study found no differences where another one did, especially if the first is designed as a replication of the second. If the findings are not confirmed, it could well be that obscure but theoretically significant variations exist in the circumstances of the two. On the other hand, it could well be that the findings of the original study were simply "sports" and lacking in theoretical implications.

The investigator accepts the latter alternative. First, Hunter's findings had been wholly unanticipated. They emerged post factum from one of the many implicit inspections he made of his data. If enough such inspections are made, one necessarily will discover differences large enough to have the probability of occuring by chance less than five times in 100 or even less than one time in 100. Statistical theory predicts that chance differences occasionally will appear as statistically significant. Note also that Hunter's five statistically significant differences (two for superintendents, two for teachers, and one for board members) were not entirely independent. Individual scores on Initiating Structure and Consideration typically show a substantial correlation. Thus, only three of the five significant differences were independent.

Second, Hunter's findings were strictly empirical. No ready explanation of them was available, either by recourse to social theory or to common sense. They could only be explained by positing relatively intricate variables which mediated between the theoretically inert fact of school system size and psychological processes of the schools' inhabitants.

A third consideration is the fact that no hint of confirmation could be discovered in the examination of two other, independent sets of data—one collected by Halpin in Ohio, the other by Evenson in Illinois. Although these studies were not closely comparable to Hunter's, they point to the fact that his findings are not capable of generalization to situations beyond that which he studied. Data from the replication study show that they cannot be generalized to closely comparable situations nor even to the same school systems three years later.

All this is not to say that Hunter's respondents behaved in an unlawful manner or the LBDQ scores they furnished were undetermined. It does suggest that whatever the various forces were which induced respondents to furnish the scores they did, these forces happened to align themselves accidentally according to school system size at the time of Hunter's investigation. In this sense, the investigator recommends that Hunter's 1958 findings be regarded as entirely fortuitous.

Part II

The LBDQ and Superintendent Behavior

Chapter 5

Introduction

When Hunter's findings failed to be confirmed, the investigator's concern that the Leader Behavior Description Questionnaire was not appropriate for describing leadership of school system superintendents came to the fore. This concern had been growing gradually. In connection with the Hunter dissertation, which the investigator directed, the substance of the LBDQ was irrelevant for Hunter's original purpose, and little critical attention was paid to it. Hunter's initial hypothesis centered on the conditions under which teachers and superintendents or board members and superintendents were more or less likely to be in agreement regarding important aspects of their work. The LBDQ was chosen as a convenient means of measuring the extent of agreement on an issue of some significance to all parties. Whether or not the LBDQ measured what it was purported to measure was not a salient question at that time.

The investigator's concerns first became manifest when Hunter discovered the relationships between LBDQ scores and school system size, direct for superintendents' self-descriptions and inverse for teachers and board members. These differences could not be explained readily in terms of differences in how the superintendent, in fact, behaved. The investigator recalled that Halpin, too, had reported absence of agreement among various school personnel in their descriptions of the leader behavior of Ohio superintendents. Following the replication study, when it was apparent that Hunter's findings no longer needed explanation, the investigator decided to bring evidence to bear upon the question of the extent to which the LBDQ scores were interpretable as veridical reports of superintendent behavior.

As the investigator prepared for the replication study, another aspect of the LBDQ forced its attention upon him--the substantive content of the items. Close inspection of the items suggested at that time that many of them referred to behavior which was not in the realm of responsibility of the public school superintendent, especially not that of the superintendent of a large school system. The analyses reported in Chapter 1 failed to show that the items were "biased" against the large system superintendent, but the investigator's suspicions lingered. What this analysis may have demonstrated was that the LBDQ was no more inappropriate for the large than for the small systems, at least of the size classes under study. The experience of eliciting responses to the LBDQ from superintendents and teachers strengthened this belief.

The Plan of Part II

These two matters--the veridicality of the LBDQ and the relevance of the LBDQ for describing leadership in the public school--are the central features of Part II of this report. Before embarking upon these matters, we will devote a chapter to certain methodological details in order to relieve the substantive chapters of their distracting burden. Chapter 6 describes the construction of two types of measure from teacher questionnaires--one of teacher contact with the superintendent and the other of teacher affect toward the superintendent and other aspects of the work situation.

Evidence regarding the extent to which the LBDQ scores may be interpreted as veridical reports about the superintendent is assembled from the various studies of the superintendency as well as from the replication study in Chapter 7. A brief chapter--Chapter 8--presents an alternative perspective from which the LBDQ scores may be viewed--as determined jointly by objective facts of the superintendent's behavior and by functional forces residing within the persons who provide the descriptions.

Chapter 9 sets forth the considerations which lead the investigator to the conviction that the LBDQ, and especially Hunter's revision of it, is unsuited for describing leadership behavior in the highly structured organization of the school. Chapter 10 summarizes the entire course of the present investigation.

Chapter 6

Measures of Teacher Affect and Superintendent Contact

In the following chapters, and especially Chapter 7, reference is made to several measures derived from teacher questionnaire responses in addition to the LBDQ measures reported in Part I. Measures of the amount and kind of contact between the teacher and his superintendent and of the teacher's affect toward four aspects of his teaching circumstance require more extensive descriptions than we wish to give in the substantive chapters, and they are reported below in turn.

Teacher Contact with the Superintendent

Following closely the procedures of Hunter[1] the investigator attempted to elicit through teacher questionnaires indications of the extent of first-hand contact between the teacher and the superintendent of his school system. Questions were asked regarding both the frequency of contacts and their nature. Frequency, alone, was insufficient; Hunter had found a number of teachers reporting high frequencies of contact but contacts of a purely formalistic character.

Questionnaire items. The two questionnaire items regarding the frequency and intimacy of teacher contacts with the superintendent follow:

---

[1] Ovid N. Hunter, Relationship between School Size and Discrepancy in Perception of the Superintendent's Behavior, Unpublished doctoral dissertation, Washington University, 1959. Pp. 52-60.

Over the last two or three years about how often have you seen the superintendent of your school to talk with him face-to-face?

Check one only.

Several times a week.

Several times a month.

Every one or two months.

Several times a year.

Not more than once or twice a year.

Have never talked with him the last two or three years.

If you have talked with your superintendent in the last two or three years, which of the following describes the nature of your contacts?

Check all phrases that apply.

Our contacts were limted to brief casual remarks on events of the day.

We have had more extended conferences on school business.

I have discussed my personal or professional problems with him.

He and I have talked as friends about our common personal interests and affairs.

I have never talked with him in the last two or three years.

Phrases concerning the nature of the contacts (if any) were chosen to fall in an order of increasing intimacy of interaction. That they do reflect such an order is strictly presumptive on the part of the investigator. Teachers were permitted to check more than one of the phrases. The large majority, however, checked but one; 17 per cent checked two and 7 per cent three of the alternatives. In scoring the intimacy question where more than one alternative was chosen, the respondent was assigned according to the most intimate of the phrases he checked.

The contact index. Responses to the two questions were combined to form a single index of amount of contact between the teacher and his superintendent. The six-point frequency scale was dichotomized to form "high" and "low" frequency categories; similarly, nature of contact was dichotomized to form "high" and "low" intimacy categories. The two sets of responses then were cross-tabulated to yield a two-by-two matrix. Each of the variables was dichotomized, however, in such a way that one of the cells lying between the High Intimacy-High Frequency and the Low Intimacy-Low Frequency cells would be vacant or approximately vacant.

Table 17, displaying the cross-tabulation of responses, shows the details of the procedure. The frequency scale was dichotomized between "several times a month" and "every one or two months," while the intimacy scale was dichotomized between "brief casual remarks" and "extended conferences." The eight respondents who did not reply to the first question regarding the frequency of their contacts were discarded. Those who failed to reply to the second question had to be inspected more carefully. Some who did not reply had indicated on the first question that they had never seen the superintendent or had seen him only once or twice a year. These 20 subjects were retained in the Low Intimacy-Low Frequency cell. Three other teachers indicated a somewhat higher frequency of contact without describing the nature of it and

Table 17

Cross-Tabulation of Intimacy and Frequency of Teacher Contact with Superintendent

Intimacy of Contact**	Frequency of Contact*							
	None	1	2	3	4	5	NA	Total
NA	13	7	1	2	-	-	6	29
None	25	2	-	-	-	-	1	28
1	1	87	60	30	32	6	1	217
2	1	25	30	17	19	6	-	98
3	1	18	26	20	28	7	-	100
4	-	21	40	39	46	44	-	190
Total	41	160	157	108	125	63	8	662

---

*1 - Not more than once, twice yearly
2 - Several times a year
3 - Every one, two months
4 - Several times a month
5 - Several times a week
NA - Not ascertained

**NA - Not ascertained
1 - Brief, casual remarks
2 - Business conferences
3 - My personal, professional problems
4 - Our common personal affairs

Note - Boxed frequencies were discarded on grounds either that insufficient information was available or that the information was patently inconsistent.

they, along with six who responded to neither question, were discarded. Two other subjects who provided clearly inconsistent information--who said they had never seen the superintendent but proceeded to describe the nature of their contacts with him--also were discarded.

The four cells of Table 17 and the weights assigned to them are summarized in the following manner:

Weight	Intimacy	Frequency	Number
2	High	High	150
1	High	Low	236
0	Low	Low	235
x	Low	High	38
x	Not ascertained		13
		Total	662

The Low Intimacy-High Frequency cell, containing but 38 cases, was designed to be virtually vacant so that the remaining three cells could be assigned an unequivocal rank order in the amount of contact displayed. In most of the analyses, cases in the fourth cell were ignored, although they were consistently inspected for unusual departures from the remainder of the cases.

For one set of analyses presented in Chapter 8 it was necessary to reconstruct the contact index as a dichotomy rather than a trichotomy. Limited to teachers in large school systems, the dichotomy was formed by assigning to a "little contact" category teachers who had never seen the superintendent and teachers whose brief, casual contacts were less frequent than every one or two months (N = 81). Teachers whose brief contacts occurred every one or two months or more frequently as well as all others with more intimate degrees of contact, regardless of the frequency, were assigned to a "much contact" category (N = 82). Four subjects were discarded on the same grounds as previously mentioned in connection with the trichotomy.

Measures of Teacher Affect

The investigator constructed four cumulative scales of teacher affect toward different aspects of their teaching situation.[2] A fifth scale to measure teachers' sense of powerlessness with regard to school affairs was constructed simultaneously for the purpose of a companion study, and its details will not be reported here.[3] The four scales of teacher affect follow.

1. Affect toward the person of the superintendent (P Scale). It was designed to measure global respect or liking for the superintendent as a leader and, thus, to supplement the brief, over-all evaluation of the superintendent's "quality as a leader" by means of a 7-point graphic rating scale. Both the P Scale and the graphic rating scale were included to check on the halo effect in the behavioral descriptions, but the P Scale was expected to be the more rigorous of

---

[2] These scales were developed before the disconfirmation of Hunter's findings and had been designed to explain the findings if they were confirmed. The rationale for choosing to measure these particular variables no longer is relevant.

[3] Gerald H. Moeller, The Relationship between Bureaucracy in School System Organization and Teachers' Sense of Power, Unpublished doctoral dissertation, Washington University, 1962.

the two. The graphic rating scale had been used for the same purpose in the early development of the LBDQ.[4]

2. Attitude toward school administrators and upper reaches of administration (A Scale). This scale attempted to differentiate between the "central office" and administration of the local school building and to measure attitude toward only the former.

3. Identification with, or loyalty to, the school system in which the teacher is employed (S Scale). This scale was designed to parallel one of the four dimensions of morale distinguished by Katz and Kahn in their report of a series of industrial studies--general satisfaction with the company.[5] Items forming the S Scale reflect a sense of pride in and loyalty to the particular school system in which the teacher is employed, quite apart from the satisfaction the teacher finds in teaching.

4. Intrinsic satisfaction with the work of teaching (T Scale). This scale, too, parallels one of the dimensions of morale discussed by Katz and Kahn--intrinsic job satisfaction. It sought to measure the enjoyment the respondent found in the work of teaching, irrespective of the system in which he worked.

---

[4] John K. Hemphill and Alvin E. Coons, "Development of the Leader Behavior Description Questionnaire," in R. M. Stogdill and A. E. Coons, (Eds.) Leader Behavior: Its Description and Measurement, Research Monograph No. 88. Columbus, Ohio: Bureau of Business Research, Ohio State University, 1957. Esp. p. 18.

[5] Daniel Katz and Robert L. Kahn, "Some Recent Findings in Human Relations Research in Industry," in G. E. Swanson, T. M. Newcomb, and E. L. Hartley, (Eds.) Readings in Social Psychology, 2nd ed.; New York: Holt, 1952. Pp. 658-659.

The construction and analysis of these scales entailed the following principal steps: (1) scale analyses and selection of items, based upon responses from a trial population of teachers; (2) modification of scoring procedures, based upon an examination of responses from teachers in the study population; and (3) an evaluative factor analysis of these responses. These steps are described in order.

Initial scale construction. The investigator and his assistant first constructed a large number of items to tap the kinds of affect to be measured. Item-by-item inspection and discussion between the researchers reduced the pool of items to 39--8 for the P Scale, 9 for the A Scale, 10 for the S Scale, and 12 for the T Scale. These 39 Likert-type items--statements with which the respondent could agree or disagree in varying degrees--were assembled in a mimeographed questionnaire for administration to a trial population of teachers. To be completed anonymously, the questionnaire was explained as a pretest for a major study to be conducted elsewhere.

The trial population was obtained in two ways. Persons attending education extension courses held by Southern Illinois University in the area surrounding Carbondale, Illinois, were asked by their instructors to respond to the questionnaire.[6] These courses were attended mainly by full-time teachers in the area school systems, and a question on the form with respect to their current occupation enabled the investigator to eliminate any respondents who were not currently in classroom teaching. By virtue of the heavy weighting of teachers from small, rural schools in returns from this source, another set of responses to the trial questionnaire was obtained from teachers in a larger suburban school system in the St. Louis area which was not to be included among the 20 systems in the study proper. Approximately 130 usable returns were available from the two sources, one-fourth of which were from the suburban system.

---

[6] We are grateful to Roland Keene, Oliver P. Kolstoe, Frank W. Lutz, George T. McClure, Benson B. Poirer, and Shelby S. Shake for their aid in getting teacher responses in the two locations.

Each of the four scales were subjected to a separate scale analysis, following the Guttman procedures for cumulative scaling.[7] Gradogram equipment was employed. This equipment consists of small multi-colored blocks which, when arrayed on a flat surface to represent the subject-by-item response matrix, receive a spindle to allow the transposition of columns and rows of the matrix.

One hundred respondents were selected randomly from the pool of 130 for each analysis. Any respondents who failed to answer one or more items of the scale under consideration were eliminated. Although the investigator attempted to retain all five response categories in the scaling, he quickly found this too complex. Hence, response categories were trichotomized and finally, in most cases, dichotomized for the scaling. Dichotomization was conducted so as to yield item marginals lying approximately between 20 and 80 per cent of the total number of responses and clustering as closely as possible around 50 per cent. The scaling followed the standard procedures which seek to minimize error through successive column and row transpositions. Where a given item contributed an excessive amount of error, it was re-dichotomized and introduced again into the matrix. Where it still contributed high error, it was removed entirely from the scale. Strenuous effort was made to retain all items of the scale's initial pool, recognizing that a too-facile elimination of items could produce a spuriously error-free scale. Such a scale would not stand up under cross-validation.

---

[7] Samuel A. Stouffer, et al., Measurement and Prediction, Princeton, N. J., Princeton University Press, 1950. Chapters 3-6.

Scales were readily achieved for three of the four measures. Except for the A Scale, little difficulty was encountered in finding at least six dichotomous items which scaled with a reproducibility coefficient of .92 or higher. (Results of the four scales analyses are shown in Table 18; the items and details of the scale analyses are presented in Tables IV and V of the Appendix.) In the case of the A Scale, however, which started with a pool of 9 items, satisfactory reproducibility (.93) could only be achieved for three items and one "contrived item."[8] One of the three items, however, was trichotomous, thereby giving an increment of strength to the scale. (Further difficulty was experienced with the A Scale in the investigation as noted below, and the investigator harbors reservations regarding its meaningfulness.)

A total of 24 items from the initial pool of 39 was found in the trial run to provide scalable measures of the four variables, and these were used in the final form of the teacher questionnaire.

---

[8] Samuel A. Stouffer, Edgar F. Borgatta, David G. Hays, and Andrew F. Henry, "A Technique for Improving Cumulative Scales," in Matilda White Riley, John W. Riley, Jr., and Jackson Toby, Sociological Studies in Scale Analysis, New Brunswick, N. J.: Rutgers University Press, 1954, pp. 372-389.

Table 18

Details of the Scale Analysis of Four Measures of Teacher Affect: Trial Run Data

	P Scale	A Scale	S Scale	T Scale
Number of Responses	95	91	92	94
Number of Items	6	5*	7	6
Number of Cutting Points	7	6	8	7
Range of Marginals, in Per Cent	24 - 83	26 - 84	21 - 83	35 - 68
Per Cent Perfect Scale Types	66	64	57	66
Coefficient of Reproducibility	.93	.93	.92	.92

---

*One contrived item and one trichotomized item.

Re-examination of the scales. After questionnaires were returned by the 662 teachers in the major study, their responses to each item were transferred to Hollerith cards before scale scores were entered for them. Some alteration in the scoring procedure was anticipated, depending upon the distribution of item marginals in this larger population of teachers. Tabulations were made of the responses to each item, and if the marginals of two items were identical, or virtually so, one of the items was dropped from the scale. Five of the 24 items were omitted from the scoring on this basis. (See Table 19.) In two instances, the marginals suggested that the order of items be transposed. This was the case for Items 5 and 23 in the P Scale and for Item 10 and the contrived item in the A Scale. Item 24 in the A Scale had to be re-dichotomized since no respondents chose the Strongly Disagree alternative which constituted the positive score under the initial dichotomization. (Inclusion of the Disagree alternative raised the marginal of this item only to 8 per cent.)

Subjects were scored on these shortened scales, then, only if they omitted no questions; if one or more responses were missing, their scores were coded Not Ascertained. A score was assigned to a subject corresponding to the perfect scale type which his scale pattern most closely approximated. If a subject's response pattern did not correspond exactly to a scale type, one of two resolutions of the issue was made. If it was possible to change one entry in his response pattern and thereby render his response pattern coincidental with a perfect scale type, he was accorded the score of that type. Occasionally, a change in either one of two entries was possible

Table 19

Reconsideration of Scale Items, Based

Upon Data from 662 Teachers

Scale	Item Number	Per Cent Positive Marginal	Disposition
P Scale	#16	27%	
	8	52	Dropped
	5	71	Order transposed with #23
	23	50	Order transposed with #5
	20	81	
	12	81	Dropped
A Scale	#24	8%	Re-dichotomized
	13	27	
	10	74	Order transposed with C*
	7 } C*	33	Order transposed with #10
	21	74	
S Scale	#22	23%	
	2	40	
	19	48	
	18	56	
	9	77	
	15	90	Dropped
	3	90	
T Scale	#11	31%	Dropped
	14	32	
	6	42	
	4	53	
	17	65	
	1	61	Dropped

---

*Contrived item formed from Items #7 and #21.

but would render the pattern coincidental with two different scale types; in this event, the subject was assigned a compromise score half way between the two alternative scale types. In those instances, however, when it would have been necessary to change two or more entries to make the response pattern conform to a perfect scale type, the respondent was coded as a Non-Scale Type. Table 20 shows the distribution of the 662 teachers over the scale scores, Non-Scale Type, and Not Ascertained codes for each of the four scales. It also shows the percentage of all scorable responses which corresponded exactly to perfect scale types.

Factor analysis of the scale items. Each of the four scales of teacher affect was developed independently of the others. Only in the first stage of scale construction, where an attempt was made to devise items unique to the variable in question, was the issue of the extent of independence or interdependence among the scales considered. While scale analysis provided some indication that the items on each scale were unidimensional, the possibility remained that some items contributed to a scale for which they were not intended or that the scales themselves were highly interdependent.

To answer these questions, a principal axis factor analysis, rotated to oblique simple structure, was conducted on the 24 items of the teacher affect scale using data from the 662 teacher responses.[9]

---

[9] The six items in the Sense of Power Scale were also included in the factor analysis, raising the number of items to 30. The analysis was conducted by the Washington University Computer Center on the IBM 650, using Calvin E. Wright's "Principal Axis Factor Analysis Program

Table 20

Number of Perfect Scale Types and Frequency Distributions of 662 Teachers over Scale Scores in Four Cumulative Scales

Scale	Score	Number of Teachers	Number Perfect	Per Cent Perfect of Total Teachers*
P Scale	5	116	110	
	4	204	155	
	3	143	113	
	2	67	50	
	1	84	72	
	Non-Scale	3	-	
	Sum	617	500	81%
	NA	45		
	Total	662		
A Scale	6	18	14	
	5	151	83	
	4	135	109	
	3	124	61	
	2	121	105	
	1	60	37	
	Non-Scale	16	-	
	Sum	625	409	66%
	NA	37		
	Total	662		
S Scale	7	116	102	
	6	107	76	
	5	86	62	
	4	52	30	
	3	142	100	
	2	78	65	
	1	33	27	
	Non-Scale	11	-	
	Sum	625	462	74%
	NA	37		
	Total	662		
T Scale	5	152	130	
	4	106	71	
	3	96	73	
	2	99	62	
	1	168	153	
	Non-Scale	5	-	
	Sum	626	489	78%
	NA	36		
	Total	662		

Four points of information bearing upon the success of scale construction were sought from the factor analysis: (1) to what extent does the number of factors extracted in the analysis correspond to the number of scales "built into" the items? (2) do the items heavily loaded on a factor correspond to the items included in a given scale? (3) do the items heavily loaded on one factor have approximately zero loadings on the other factors? and (4) how great is the correlation between the factors?

Six factors were extracted by the analysis before the residual matrix entries were sufficiently small to register a program cut-off. Subsequent interpretation of the rotated matrix indicated that the first factor to be extracted corresponded to the P Scale. Other factors appeared in this order: T Scale, S Scale, Sense of Power Scale, A Scale, and an uninterpreted Factor VI. The number of factors, then, was one in excess of the number of scales "built into" the items. On this criterion, scale construction was successful.

With respect to the correspondence between scale items and factor loadings, scale construction also was unequivocally successful. For example, the most heavily loaded items on Factor I, ranging from .53 to -.67, were precisely the items designed for the P Scale. No other item loadings on this factor even approached .53; the highest in absolute value was one item with a loading of -.28. The same was

---

for the IBM Type 650," University of Washington, April, 1957, and Mary R. Nickles and Thomas A.Keenan's "IBM 650 Program for the Pinzka-Saunders Solution for Oblique Rotation to Simple Structure," University of Rochester Computing Center, n.d. Appreciation is hereby extended to Arch H. Turner, Jr., of the Washington University Computer Center who carried out the machine operations.

even more dramatically true for the other three factors of relevance to the present discussion. (Item loadings for all six factors are displayed in Table VI of the Appendix.) Furthermore, items loaded heavily on one factor universally had low loadings on the three other factors corresponding to the scales. The pattern was so clear that the investigator had no difficulty in matching items to scales simply from inspection of the factor loadings.

Finally, the angle separation between the obliquely rotated factors provided information regarding their intercorrelations. Table 21 shows the cosines of the factor angles, or the coefficients of correlation between the factors. Factors for the P and A Scales, the A and S Scales, and the A and T Scales were virtually orthogonal (correlations of .07, .01, and .03, respectively). The P and S Scale factors, on the other hand, were substantially correlated (.42), while the remaining two factors--for the P and T Scales and the S and T Scales--were modestly correlated (.17 and .25, respectively). Thus, limited success was realized in constructing empirically independent scales of teacher affect. Even the highest of the correlations was not so high as to suggest that the two scales measured the same variable in alternative ways. A coefficient of correlation of .42 leaves approximately 82 per cent of the variance unexplained.

In sum, the evidence from both the Guttman scale analyses and the factor analysis points strongly to the existence of four unidimensional, cumulative scales. The most reasonable inference regarding their psychological meaning is that they measure the aspect of teacher

Table 21

Cosines of Angles between Factors, Corresponding to Five Cumulative Scales, from Oblique Rotation of Principal Axis Factor Analysis

	P Scale	A Scale	S Scale	T Scale
A Scale	.067			
S Scale	.420	.013		
T Scale	.169	.026	.253	
Sense of Power Scale	.361	.199	.169	.120

affect which they were constructed to measure. No concentrated effort was made in this study, however, to establish the validity of the measures.

The A Scale seems to be the weakest of the four scales from the standpoint of the difficulties encountered in its development. In the initial scale analysis, satisfactory scalability was achieved only after all but five of the items had been eliminated and two of these combined to form a contrived item. Two revisions in scoring of the scale were necessary from an examination of the responses of the new population of 662 teachers: one item was re-dichotomized and the order of two other items was reversed. Even with the re-dichotomization of the one item, the marginal for that item was extremely low, a circumstance which would articially inflate an estimate of scale reproducibility. In the factor analysis, the factor corresponding to the A Scale was the next to the last to appear (the last being the uninterpreted Factor VI), and the factor loadings on two of the five A Scale items, although discriminably higher than those on any item alien to the A Scale in the factor, were comparatively low. The loadings of the two items--.26 and -.31--were less than the loadings of the keyed items in the other three scales on their respective factors, all of which were .45 or higher in absolute value. Finally, the various substantive analyses conducted in connection with the LBDQ failed to demonstrate many psychologically meaningful correlates of teacher affect toward upper-level administrators and administration. For these reasons, no further reference to the A Scale is made in this report.

Chapter 7

The LBDQ as a Veridical Report on Superintendent Behavior

Examination of Table III in the Appendix illustrates the marked variation in scores yielded by the LBDQ, whether the forms are completed by the superintendents' associates or by themselves. What accounts for these gross variations? The most parsimonious explanation is that the scores reflect differences in the actual leadership styles of the superintendents--that they represent, within the margin of measurement error, veridical descriptions of leader behavior. This is, after all, what the LBDQ was devised to measure.

Unfortunately, the data of the replication study and of Hunter's study as well as those of previous investigations are not fully consistent with a "veridicality" interpretation. The present chapter will assemble the relevant evidence from published studies on the school superintendency and will present results of the investigator's analyses of data from the replication study.

One convincing way to demonstrate that the LBDQ scores reflect directly and veridically the manifest behavior of superintendents is to check the scores against an independent criterion of leader behavior, such as one provided through systematic observation by trained observors. The one group of investigators who has attempted to observe superintendent behavior encountered difficulties in identifying behavioral acts which could unequivocally be classified as Initiating Structure or Consideration.[1] No efforts to correlate LBDQ scores with a criterion measure are available in the superintendency literature.

In the absence of such evidence, support for the veridicality interpretation must be sought in other ways, the more prominent of which are:

1. Evidence that the LBDQ scores relate to other variables in ways which could be predicted on theoretical grounds (construct validity).

2. Evidence that a respondent is able to discriminate among various kinds of behavior of a given leader.

3. Evidence that a respondent is able to discriminate between the behavior of several leaders.

4. Evidence that different respondents agree on the behavior of a given leader.

A fifth kind of evidence sometimes adduced to support the validity of the LBDQ would not be strongly convincing at best--evidence that different groups of respondents, each describing different leaders, generate significantly different sets of scores. Such evidence typically is provided by means of analyses of variance conducted across school systems. Significant F ratios, however, may equally as well imply that respondents differed from one school system to another as that the behavior of their superintendents differed. Suffice it to say that two investigators (Halpin and Hunter) found significant F ratios on the tests which they ran and one (Seeman) did not.[2]

---

[1] James M. Lipham, "Initiating Structure and Consideration," [illegible] Observation of Administrator Behavior, Chicago: Midwest Administration Center, University of Chicago, 1959. Pp. 27-68.

While other kinds of evidence might be sought in support of the veridicality interpretation, the four listed above cover and even extend beyond the kinds available in the published literature on the school superintendent. Discussion in the present chapter will be limited to the four.

Relationship with Other Variables

It should be noted immediately that none of the studies of superintendent leadership proposed theoretical relationships between leader behavior and other intimately connected variables which, in their testing, could provide evidence of construct validity. Where empirical relationships were examined, either the other variables were fortuitous ones, as in the case of school system size, or they were irrelevant to superintendent behavior, as in the case of attributes of the teacher respondents. In fact, it was the very unreasonableness of Hunter's findings regarding the LBDQ and school system size which gave rise to the replication study. One would normally expect, albeit on theoretically unpretentious grounds, that superintendents in larger systems would manifest greater leadership competence, hence score higher on the LBDQ, than those in the smaller systems. Hunter discovered just the reverse, at least through the reports of teachers and school board members, and, as we have described in Part I, even the reasonable relationships with size could not be consistently found in other studies.

---

[2] Andrew W. Halpin, Leadership Behavior of School Superintendents, School Community Development Study Monograph No. 4, Columbus, Ohio: Ohio State University Press, 1957; Ovid N. Hunter, Relationship between School Size and Discrepancy in Perception of the Superintendent's Behavior, Unpublished doctoral dissertation, Washington University, 1959; Melvin Seeman, "A Comparison of General and Specific Leader Behavior Descriptions," in Ralph M. Stogdill and Alvin E. Coons, (Eds.) Leader Behavior: Its Description and Measurement, Research Monograph No. 88, Columbus, Ohio: Bureau of Business Research, Ohio State University, 1957. Esp. p. 96.

Two studies, the replication study included, examined relationships between teacher reports of the superintendent's leader behavior and outside variables pertaining to attributes of the reporting teachers. Some of these will be discussed now and others reserved for the next section dealing with the issue of halo. Were relationships found between LBDQ scores and attributes of those furnishing the scores, far from supporting the veridicality interpretation they would argue against it. They would suggest that the LBDQ reports reflected characteristics of the reporters, not the leader behavior which ostensibly was being reported.

In an investigation of 71 elementary school teachers reporting on their 12 superintendents, Seeman[3] showed a general absence of relationship between Initiating Structure and Consideration scores, on the one hand, and five personal history variables of age, salary, years of teaching experience, length of time in present school system, and years served under the present superintendent. Only the relationship (-.39) between Initiating Structure and years of teaching experience was statistically significant. Neither LBDQ dimension correlated significantly with a measure of the teachers' personal job satisfaction.

---

[3] Seeman, op. cit.

The replication study obtained two measures corresponding to different aspects of job satisfaction--the teacher's identification with the school system (S Scale) and the extent to which the teacher finds teaching an intrinsically satisfying kind of work (T Scale). (Chapter 6 describes the scales and their construction in detail.) Table 22 displays the correlations between these measures and the two LBDQ dimensions for the "regular" and "personal" forms of the LBDQ. Of the four relationships in the "regular" form, only that between Consideration and the S Scale was statistically significant. It's magnitude (.19) is not sufficiently large to account for much of the variance in the Consideration dimension. It is interesting to observe in the "personal" forms, though, the presence of two substantial relationships, between the S Scale and Initiating Structure and Consideration scores respectively. It appears that the "personal" form was more predictive of one aspect of job satisfaction, at least, than the "regular" form.

While a theoretical case could be advanced regarding a direct relationship between the extent to which superintendents initiate structure and are considerate and the extent of their teachers' job satisfaction, the foregoing data must be rejected as providing evidence of construct validity on several grounds. First, on the operational

Table 22

Correlations of Teacher LBDQ Responses, "Regular" and "Personal" Forms, with Two Measures of Teacher Morale

	Morale Measures	
LBDQ	Identification with System (S Scale)	Satisfaction with Teaching (T Scale)
"Regular" Form		
Initiating Structure	.05	.02
Consideration	.19**	.05
"Personal" Form		
Initiating Structure	.29**	.18**
Consideration	.38**	.22**

**$p < .01$

Note. The number of cases on which the correlations are based vary from 221 to 254.

level the empirical relationships were found between teacher reports of their job satisfaction and teacher reports of leader behavior. Hence, an equally tenable theoretical case could be made that the teachers' level of job satisfaction colored their reports on the superintendent. As a matter of fact, this was precisely the reason Seeman gave for his investigation. He wished to discover the extent to which LBDQ scores were contaminated by personal factors, including job satisfaction of the respondents. The same emprical data can hardly be used at one time to support the validity of the LBDQ and at another time to challenge it.

Second, both components of the job satisfaction measure used in the replication study are significantly correlated with the two measures of generalized feelings about the superintendent which we use in the next section to test for halo in the LBDQ.[4] Affect toward the superintendent, job satisfaction (especially the S Scale component), and teachers' LBDQ reports (especially on Consideration) are "all of a piece"--all pointing to a general attribute of responding teachers. Third, if we disregard the preceding reasons for rejecting the relationships found in the replication study between job satisfaction and LBDQ

---

[4] The correlations are tabled below. Based upon N's of approximately 550, all are statistically significant beyond the 1 per cent level.

	S Scale	T Scale
Graphic rating scale	.50	.21
P Scale	.47	.24

scores as evidence supporting construct validity, we can nevertheless point to the fact that Seeman found no such relationship in his study. Thus, the evidence would be mixed.

In another study Seeman investigated relationships between leadership style and superintendents' social mobility and attitudes toward mobility.[5] He posited no precise relationships with leader behavior upon which construct validity could be tested--at least, none so specific that non-confirming evidence would have led him to question the LBDQ's validity. The principal burden of Seeman's article was to propose that conditions of the superintendents' social mobility led them to distort their self-descriptions on the LBDQ. His analyses on the point were exploratory in nature rather than definitive tests (and they entailed the assumption that the reports on the superintendent by teachers and school board members were veridical), but if perceptual distortion does occur as he suggests, superintendents' self-descriptions could not be veridical.

In sum, there is no firm evidence regarding construct validity available in the superintendency studies. The occasional relationships found between LBDQ scores and other variables, particularly variables in teachers who furnish the scores, suggest that the scores are not purely reflections of objective fact regarding the superintendent.

---

[5] Melvin Seeman, "Social Mobility and Administrative Behavior," American Sociological Review, 1958, 23, 633-642.

Discriminations in the Behavior of a Leader

To regard the LBDQ as a veridical description of leader behavior, it must be shown that respondents are able to discriminate between instances of different behavior of a given leader. High correlations between responses to all items of behavior, whatever their content, would cast serious doubt upon the validity of the measure. In a form such as the LBDQ, with two sets of 15 items measuring two kinds of leader behavior, the matter of item correlations separates into two issues, only one of which bears upon veridicality. High correlations should obtain within each set of items, but the two sets should be reasonably independent of one another. These are issues of internal consistency and dimension independence, respectively.

Internal consistency of items was a prime consideration throughout the developmental research on the LBDQ at Ohio State University, and both item analyses and factor analyses provided the criteria by which internally consistent items were retained in the instrument. No further tests of internal consistency have been made in the school superintendency studies, and inasmuch as the issue does not impinge directly upon the veridicality interpretation it was not pursued in data of the replication study.

Ability of respondents to discriminate among different kinds of behavior of a leader may be approached operationally in two ways. One way, mentioned above, is through tests of dimension independence--the display of correlation coefficients between scores on each dimension. The other is through tests for halo. Halo refers to the respondent's inclination to confound his presumably objective descriptions of behavior with his evaluative judgments of the person who manifests the behavior, and a common procedure used to conduct the tests entails constructing direct measures of evaluative judgments or generalized affect toward the leader. Correlations of these with dimension scores or item responses on the behavioral description instrument suggest the extent to which halo is in operation. Let us now turn to a discussion of dimension independence and tests for halo in the LBDQ in order.

Independence of dimensions. With respect to dimension independence, we should note that investigators during the developmental period of the LBDQ were not fully successful in achieving independent measures of Initiating Structure and Consideration. Correlations of scores between the two dimensions, based upon various LBDQ forms and various samples, ranged from an atypically low .28 to a high of .56 in the early investigations. Normally the dimension correlations ran from .38 to .52.[6] One investigator, in devising a modified form for measuring supervisory behavior in industry, made an assiduous effort to "uncorrelate" the dimensions by selecting for his scale only those items which loaded heavily on one factor and close to zero on the other.[7] In this case, the investigator was successful: dimension correlateions of three samples were -.02, -.05, and -.33.

---

[6] Ralph M. Stogdill and Alvin E. Coons, (Eds.) Leader Behavior: Its Description and Measurement, Research Monograph No. 88, Columbus, Ohio: Bureau of Business Research, Ohio State University, 1957. Pp. 48, 55, 92, 93, 105.

[7] Edwin A. Fleishman, "A Leader Behavior Description for Industry," in ibid., pp. 103-119.

Dimension independence in Halpin's Ohio study varied with the group of respondents for whom the correlations were calculated. Initiating Structure and Consideration scores correlated .18 for staff respondents and .61 for school board members.[8] Apparently, the subordinates who worked closely with the superintendent were better able to discriminate between the two kinds of leader behavior than the more remote board members.

Hunter did not compute dimension correlations in his investigation, but they have been calculated in the replication study for responses of administrative staff members and teachers, and in the latter case for both the "personal" and "regular" forms of the LBDQ. The coefficients follow.

	Correlation	Number
Administrative staff	.71	102
Teachers ("regular" form)	.24	228
Teachers ("personal" form)	.21	216

All of the correlations differ significantly from zero beyond the 1 per cent level. The absence of independence is most striking for administrative staff members. (It should be recalled that this category is not identical to Halpin's "staff" and that item rewording in the LBDQ form of the replication study made the items more relevant for teachers than for staff members; see Chapter 1.) For teachers, the Initiating Structure and Consideration scores are only modestly correlated, and the "personal" form shows a slightly lower relationship than the "regular" form.

---

[8] Halpin, *op. cit.*, p. 41 (in footnote to his Table V).

It appears from the foregoing that the difficulty the investigators encountered in the developmental research in assuring respondent discrimination between the two kinds of leader behavior also was encountered in the studies of superintendent leadership. Some groups of respondents were able to make the discrimination and others were not.

Halo in the LBDQ. When raters respond to descriptive statements in terms of their generalized impressions and feelings about the person they are rating, their ability to discriminate among different aspects of the person suffers. The result is that scores reflect the global feelings of the rater, not the objective attributes of the person rated.

Halo, like dimension interdependence, was a pervasive problem in the initial studies of the LBDQ. Hemphill and Coons reported that 90 of the 150 items in the initial instrument correlated in excess of .30 with a single evaluation of the leaders' over-all "leadership quality."[9] A subsequent study found both Initiating Structure and Consideration scores substantially correlated with a nine-item evaluative measure (r = .36 and .68, respectively).[10] Neither Halpin nor Hunter furnished similar estimates of the intrusion of evaluative judgments in the LBDQ responses.

---

[9] John K. Hemphill and Alvin E. Coons, "Development of the Leader Behavior Description Questionnaire," in Stogdill and Coons, (Eds.) op. cit., pp. 18-19.

[10] Melvin Seeman, "A Comparison . . .", op. cit., p. 96.

In the replication study, the investigator included a measure of respondents' over-all evaluation of superintendents' leadership in the questionnaires to teachers identical to the form used by Hemphill and Coons. It was a graphic rating scale with numbers from 1 to 7 spaced at intervals along a line segment, with the ends of the line described as "poor" and "perfect." Respondents were asked to circle the number showing their estimate of the superintendent's "quality as a leader." In addition, the investigator constructed a multi-item cumulative scale to measure respondents' level of affect toward the person of the superintendent--the P Scale described in detail in Chapter 6. These two measures were regarded as alternative ways of assessing global feelings about the superintendent. The graphic rating scale cast the evaluation specifically in the context of leadership quality, while the P Scale employed, in addition to leadership evaluation items, several items regarding personal attraction to or liking for the superintendent. As expected, the correlation between the two measures was high, .73.

Considering the unusually low inter-dimension correlations for teachers in the replication study, the evaluative measures would not be expected to correlate highly with both Initiating Structure and Consideration. The coefficients shown in Table 23 indicate that substantial relationships exist in the Consideration dimension, less substantial relationships in the Initiating Structure dimension of the LBDQ. This, incidentally, was the same pattern which Seeman found

Table 23

Correlations of Teacher LBDQ Responses, "Regular" Form, with Two Teacher Evaluations of Superintendents

	LBDQ		
Evaluation of Superintendent	Initiating Structure	Consideration	Significance Level of Difference between I.S. and Cons.
Graphic rating scale	.26**	.59**	$p < .01$
Affect toward Superintendent - P Scale	.18*	.55**	$p < .01$

---

*$p < .05$

**$p < .01$

Note. The number of cases on which the correlations are based vary from 227 to 249. For the method of computing significance levels of the difference between the Initiating Structure and Consideration correlation coefficients, see Footnote 12 in the text.

between the LBDQ dimensions and his nine-item evaluation measures.[11] All of the coefficients in Table 23 differ significantly from zero beyond the 5 per cent level and three of them beyond the 1 per cent level. Moreover, the correlations for the Initiating Structure and the Consideration scores differ significanlty from one another beyond the 1 per cent level.[12] That is to say, there is significantly greater contamination by halo in the Consideration than in the Initiating Structure measure.

The finding that general evaluations of the superintendent enter importantly into descriptions of Consideration behavior but less into descriptions of Initiating Structure behavior raises the question as to why the difference exists. There is no reason to believe that the items in the Consideration dimension are more vague or indeterminant for responding teachers than those constituting the Initiating Structure dimension. A more likely reason is that teachers are more sensitive to the regard which superintendents show for them as individuals than to the task-oriented aspects of the superintendents' performance. They may invest greater affect in the considerateness of leaders then in their organizational capacities. Halpin pointed to precisely this tendency among educators in his comparison of the leadership ideologies of aircraft commanders and school administrators.[13] If this reasoning is adopted, it argues against the halo hypothesis. It holds that teachers' general evaluations of the superintendent are the consequence of his considerateness not antecedent conditions which bias their descriptions of his considerateness. This issue, however, cannot be settled with the data at hand.

---

[11] Loc. cit.

[12] The statistical test appropriate for the null hypothesis that two correlations do not differ when the correlations are computed for the same population is described in Helen M. Walker and Joseph Lev, Statistical Inference, New York: Holt, 1953. Pp. 256-257.

Another possible reason for the lower relationship of Initiating Structure scores with the evaluations of the superintendent resides in the character of the LBDQ items. Three of the 15 items in the Initiating Structure scale require respondents to endorse socially undesirable statements about the superintendent in order to describe him as high on that dimension. No such items exist in the Consideration scale. The three items follow:

Item 6. Rules teachers with an iron hand.

Item 8. Criticizes poor work by teachers.

Item 9. Speaks to teachers in a manner not to be questioned.

Inspection of response distributions on the items indicates that they are in fact, infrequently endorsed by the teachers. This is especially the case for the first two, where only 10 per cent of the teachers reported their superintendents as "always" or "often" behaving in these ways. On item 9, 33 per cent endorsed the "always" or "often" categories.

That these three items reduce the relationship of the Initiating Structure scores with the evaluation measures is clear. A distinct negative relationship between Items 6 and 8 and the graphic rating scale is depicted in Figure 2. A curvilinear relationship is indicated for Item 9. Plots for the other 12 items in the Initiating Structure scale uniformly show positive relationships with the evaluative measure.

---

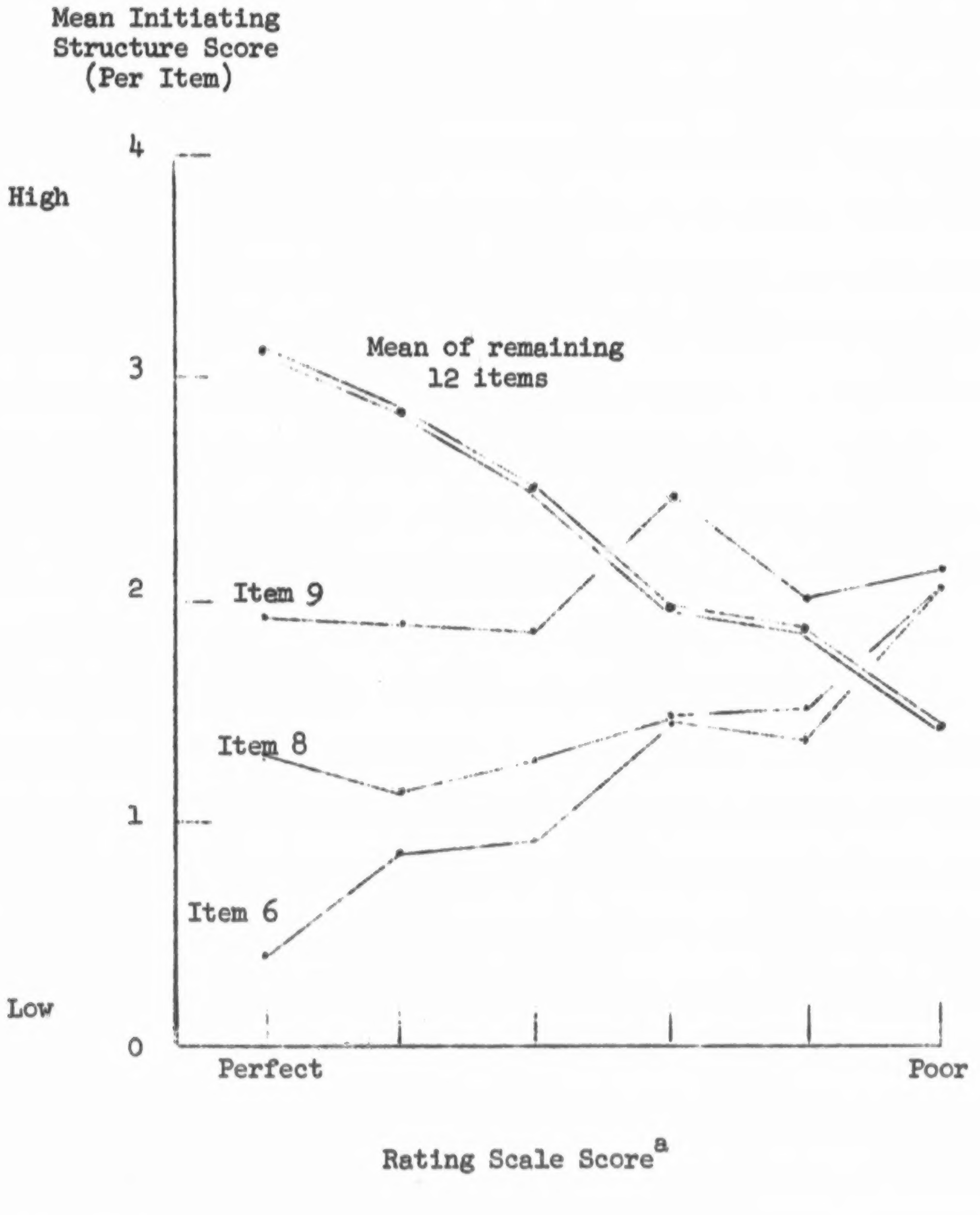
[13] Andrew W. Halpin, "The Leader Behavior and Leadership Ideology of Educational Administrators and Aircraft Commanders," Harvard Educational Review, 1955, 25, 18-32.

These observations suggest, then, that the Consideration dimension items are more heavily or more consistently imbued with social desirability than those items of the Initiating Structure dimension. In a sense, it is easier for respondents to select the Consideration items which, at the same time, render a high dimension score and express their positive affect toward the superintendent.

Halo in the "personal" form of the LBDQ. The "personal" form of the LBDQ, completed by a random half of the teachers in the replication study, asked respondents to describe how the superintendent behaved toward them personally rather than toward teachers in general. This modification supposedly made the questionnaire items less ambiguous and less subject to alternative interpretations by the respondents. It would seem that the greater objectivity of the "personal" form would reduce the extent to which generalized evaluations of the superintendent colored teachers' descriptions of his behavior.

The correlations of Initiating Structure and Consideration with the two measures of teachers' evaluations of their superintendents, in Table 24, display about the same pattern for the "personal" form that was found for the "regular" form, in Table 23. Significant relationships between the LBDQ scores and evaluative judgments still occur in three of the four cases. As in the case of the "regular"

Figure 2. Mean Item Responses on Three Selected Initiating Structure Items ("Regular" Form) for Teachers Who Evaluated Superintendents Variously on the 7-point Rating Scale



---

[a] Scores of 6 and 7 (poor) are combined to augment the number of cases.

Table 24

Correlations of Teacher LBDQ Responses, "Personal" Form, with Two Teacher Evaluations of Superintendents

	LBDQ		
Evaluation of Superintendent	Initiating Structure	Consideration	Significance Level of Difference between I.S. and Cons.
Graphic rating scale	.20**	.35**	n.s.
Affect toward superintendent - P Scale	.13	.51**	$p < .01$

**$p < .01$

Note. The number of cases on which the correlations are based vary from 203 to 233. For the method of computing significance levels of the difference between the Initiating Structure and Consideration correlation coefficients, see Footnote 12 in the text.

form, less halo exists in the Initiating Structure dimension than in the Consideration dimension scores. (The differences between Initiating Structure and Consideration correlations with the evaluative measures, however, is statistically significant only in the case of the P Scale measure.)[14]

Comparing the success of the "personal" and "regular" forms of the LBDQ in avoiding halo, it can be observed that the relationships with the evaluative measures in Table 24 are consistently lower than the relationships for the "regular" form responses in Table 23. Most of the differences between the respective coefficients are small and statistically insignificant, but the Initiating Structure - P Scale correlations of .59 for the "regular" form and .35 for the "personal" form differ significantly from one another beyond the 1 per cent level.[15] While generalized evaluations still enter importantly into the LBDQ scores attributed by teachers to their superintendents, the "personal" form is somewhat less afflicted by halo than the "regular" form.

These findings are noteworthy in connection with the question raised in Chapter 3 of the report, where evidence was presented to show that a difference in Initiating Structure scores associated with school system size existed in teacher responses on the "regular" form but not on the "personal" form. At that point we offered the possibility that the more ambiguous "regular" form may have permitted teachers to distort their descriptions of the large system superintendents in accordance with the greater status of officering a large school system. Although the present analyses do not answer the question directly, they do suggest that the "regular" form has greater potentiality for distortion, from whatever source it may come, than the "personal" form. This is especially true in the Initiating Structure dimension.

---

[14] See footnote 12.

[15] The conventional z transformation was employed to test the null hypothesis that the correlation coefficients in the two independent populations did not differ from one another beyond the bounds of chance. Walker and Lev, op. cit., pp. 255-256.

In sum, the evidence is not strong concerning the ability of respondents to discriminate among the various behaviors of a given superintendent. Dimension correlations occasionally are extremely high, and halo was shown to intrude upon objective descriptions of the superintendent, especially of his Consideration behavior.

We must point out again that an alternative interpretation is possible in the results of the tests for halo employed in this and in earlier studies. A high correlation between judgments of a superintendent's leadership quality and descriptions of his leader behavior does not, *ipso facto*, mean that the global judgments operate to bias the behavioral descriptions. Direction of causality cannot be inferred from a simple correlation coefficient. One might argue that teachers' observations--veridical observations--of how the superintendent behaves may well be the factor which shapes their judgments of his quality as a leader. We will return to this possibility in the following chapter.

Discrimination between Several Leaders

Veridicality of the LBDQ report would be questionable were respondents unable to discriminate between the behavior of two or more leaders who reasonably could be assumed to differ in their leadership styles. The superintendency studies do not furnish information on this since school personnel supplied descriptions for only their own superintendents. A corallary on which there is information, however, holds that veridicality also would be questionable if respondents described differently the behavior of superintendents who could be assumed not to differ markedly in their styles.

In Chapter 3 we reported analyses of teacher descriptions of their superintendents and superintendent self-descriptions supplied in 1958 and again in 1961 in the replication study. It is unlikely that many of the teacher respondents, especially in the large school systems, were selected in both samples and thereby furnished two descriptions; the best that can be said for them is that the samples yielded reasonably good estimates of the mean scores for the population of teachers in each system at the two points in time. In any event, comparison of the scores indicated that for neither teacher reports nor superintendent self-descriptions was there more than a chance relationship from 1958 to 1961.

It is conceivable that the leadership styles of the superintendents in fact changed in the course of the three years and that the absence of test-retest correlations simply reflected this variability in behavior. Upon closer investigation of the scores in each system, however, we showed in Chapter 3 that the superintendent and the teaching staff did not agree on the direction of any changes which might have occurred. The data even suggested that they were inclined to report changes in opposite directions. These analyses, as limited as they may be, defy explanation by means of the assumption that the LBDQ is a veridical description of leader behavior.

Agreement among Respondents on a Given Leader

The most abundant kind of evidence regarding the veridicality interpretation of LBDQ scores is on the question of the extent to which different respondents agree in their descriptions of a given superintendent. It is reasonable to suppose that, were the reports of respondents governed only by the objective fact of the superintendent's behavior, their reports would be virtually identical, limited only by measurement error. This implies a) that the superintendent behaves consistently toward the different respondents and b) that the respondents have equal access to information about his behavior.

The possibility that the first of the two implications does not hold was raised by Halpin in his attempt to explain why classes of respondents in his study did not agree in their descriptions of the superintendent.

> Although the staff and the board members agree among themselves as a group in their description of the superintendent's Consideration, they do not agree with each other. Likewise, the staff and the board members agree among themselves as a group in their description of the superintendent's

> Initiating Structure behavior but do not agree with each other. This suggests that the superintendent adopts different role behavior in dealing with each of these groups.[16]

In other words, Halpin asks us to assume that the respondents' reports were veridical but that the superintendents' behavior was inconsistent. The difficulty with this reasoning is that it fails to instruct us in advance as to which respondents should and which should not agree and as to the direction of the disagreement. Under what circumstances is absence of agreement evidence in support of veridicality and under what circumstances is it evidence against veridicality? By indirection he suggests that agreement within a category of respondent (school boards, staff) can be expected but that agreement between categories need not be expected. (The direction of the disagreement between categories comes only from his empirical findings, which show school boards describe the superintendent higher than staff on both Initiating Structure and Consideration.) Halpin's *post factum* interpretation preserves the assumption of veridicality while our intent is to test the assumption, so we shall hold open the issue of whether within-category or between-category agreement is the more significant evidence regarding veridicality.[17]

---

[16] Halpin, *Leadership Behavior of School* . . . , *op. cit.*, p. 55. Italics added.

[17] Within-category and between-category agreement corresponds to the distinction drawn between intra-position consensus and inter-position consensus in Neal Gross, Ward S. Mason, and Alexander W. McEachern, *Explorations in Role Analysis: Studies of the School Superintendency Role*, New York: Wiley, 1958. Pp. 104-115.

As indicated above, the data on agreement in LBDQ scores can be viewed in two ways--in the extent of agreement between different groups of reporters and in the extent of agreement within each group of reporters. Agreement between groups is readily expressed by correlation coefficients, but the expression of the extent of within-category agreement poses a problem. The variance of scores around the group mean is an appropriate measure of within-group agreement (the greater the aggreement the less the variance), but no standard exists by which a given variance figure can be evaluated. How small a variance represents "agreement" among respondents?

The investigator has met the problem of evaluating the variance figures by capitalizing on the second implication, b, stated in the opening paragraph of this section. The reasoning is straight-forward. If the reports of respondents are governed only by the objective fact of the superintendent's behavior, then those respondents who have the greatest access to information about this behavior should agree most closely on it. Thus, the extent of agreement (the variances) can be compared among categories of respondent who vary in their access to the superintendent. The procedure will be described in the second section below.

Between-group agreement. Both the Halpin and the Hunter studies of school superintendents demonstrated that different categories of respondent did not often agree in their descriptions of the superintendents' leader behavior. In the Ohio data, correlations between reports furnished by board members, by the staff, and by the superintendents did not regularly depart from zero. Only in the superintendent-staff comparison on Initiating Structure was the relationship greater than chance.[18]

In Hunter's investigation, superintendents disagreed with teachers and board members, in one direction in the large school systems and in the other direction in the small systems. The strongly negative correlations between the superintendent self-descriptions and the descriptions provided by their associates, however, was accompanied by agreement between the associates, at least in the case of Initiating Structure (.81). The correlation between teachers and board members on Consideration did not achieve statistical significance at the 5 per cent level.[19]

Hunter's findings of high agreement between teachers and school board members cast a shadow over Halpin's suggestion that superintendents are inclined to differentiate their role behavior toward superiors and subordinates. It leaves us with the question as to why Hunter's superintendents did not behave in a differentiated way. Or perhaps they did, but the respondents failed to reflect it in their reports. If this were the case, the reports could not be regarded as veridical.

---

[18] Halpin, Leadership Behavior of School . . . , op. cit., p. 50.

[19] The investigator computed the rank difference correlations from Hunter's data over his 16 school systems, using school means in the case of teachers and board members. Superintendent-teacher, superintendent-board, and teacher-board correlations on Initiating Structure were -.78, -.46, and .81 and on Consideration -.46, -.47, and .46 respectively. Only the two highest were statistically significant, and those beyond the 1 per cent level. Levels of statistical significance for r_d were established by reference to Table G in Merle W. Tate and Richard C. Clelland, Nonparametric and Shortcut Statistics, Danville, Illinois: Interstate Publishers, 1957. P. 132.

(We should note that the mean scores of Hunter's teachers were generally lower, in keeping with Halpin's empirical observation, than the means for board members.)

The replication study furnished inconsistent evidence on between-group agreement. Correlations between teachers, administrative staff members, and superintendents are displayed in Table 25. Rank difference coefficients were used because of heterogeneity of variance in the scores of teachers and administrative staff. The statistically significant correlations of .67 and .62 between teachers and administrative staff members indicates that they tended to agree in the amount of Initiating Structure and Consideration behavior they saw in their superintendents. Neither group, however, agreed beyond chance with their superintendents. Thus, all three studies suggest that superintendents' self-descriptions are undependable, at least in the sense that the LBDQ scores are usually uncorrelated with scores furnished by their associates: one correlation was significantly positive, another significantly negative, and six insignificantly different from zero.

Within-group agreement. To what extent do groups of respondents agree among themselves in describing the leader behavior of school superintendents? Both Halpin and Hunter noted statistically significant differences in the amount of agreement (or, inversely, in the extent of within-group variances in scores) from one school system to the next among certain of their categores of respondents for

Table 25

Rank-difference Correlations between Superintendent, Teachers, and Administrative Staff on Initiating Structure and Consideration

	Initiating Structure		Consideration	
	Staff	Teacher	Staff	Teacher
Superintendent[a]	.22	.32	.01	-.05
Staff[b]		.67*		.62*

---

*$p < .05$

[a] $N = 19$

[b] $N = 20$

either Initiating Structure scores, Consideration scores, or both. The same was true in the replication study. That the variances are inhomogeneous across school systems, however, is not a strong case one way or another with respect to validity. The best that can be said on this basis is that the LBDQ is less veridical--that there is less agreement--in some settings than in others.

Halpin, therefore, went on to compute values of the unbiased correlation ratio (Epsilon) for staff and board members in his 50 systems. Epsilon values ranged between .44 and .63 on the two LBDQ dimensions for the two groups, and all were significant by the F test well beyond the 1 per cent level.[20] But Epsilon is not an entirely satisfactory measure of within-group agreement. Its value (one minus the ratio of within-group variances to the total sum of squares about the grand mean) is a function of the variation in the elevation of mean scores from one school system to the next. For a given magnitude of the within-group variance, Epsilon would be zero if the school system means were identical and would be close to 1.00 if the system means differed greatly from one another. In other words, the standard against which within-group agreement is compared by Epsilon is a highly relative one.

In the replication study, a more suitable approach to the comparison of variances was devised by invoking the assumption that personnel who have the most direct access to relevant information about the superintendent's behavior should agree more in their characterizations of his behavior than personnel who have little or no access to such information. Within each school system, teachers and administrative staff members were separated into four groups constituting different degrees of contact with the superintendent. Since administrative staff members were nominated by superintendents as the persons with whom they had the closest working relationship, they constituted the first "contact group." Teachers were grouped according to their contact index scores, derived from their questionnaire replies concerning the frequency and nature of their interaction with the superintendent. (See Chapter 6.) Those who fell in the High Intimacy-High Frequency category of the contact index were considered to have the next greatest access to the superintendent after members of the administrative staff. Teachers who fell in the High Intimacy-Low Frequency category were placed next, and those in the Low Intimacy-Low Frequency category were considered to have the least access to information regarding the superintendent's leader behavior. Many of the teachers in the last category said they had never seen the superintendent to talk with him in the preceding three years.

---

[20] Halpin, Leadership Behavior of School . . . , op. cit., pp. 46-47.

The variances of Initiating Structure and Consideration scores ("regular" form only) were computed for each group of personnel in each school system. If, however, a school system had fewer than three teachers in any given category, as sometimes happened, no computations were made and the datum for that group was omitted. Only three school systems had sufficient numbers of personnel to provide variance data in all four contact groups. In most systems, only three variance figures could be computed.

Each adjacent pair of variances available within a school system were compared, and a confirming instance was tabulated if the variance for the group with the closer contact with the superintendent were the smaller of the pair. (Comparison of all pairs within a system, whether adjacent of not, would be inappropriate since only the adjacent ones are independent of one another.) Thirty-four comparisons could be made for Initiating Structure variances and 38 for Consideration variances in the twenty school systems. The veridicality interpretation would be supported if the confirming instances outnumbered significantly the disconfirming instances.

Tabulations of confirming instances in Initiating Structure scores indicated that in 20 of the 34 comparisons of variances the group of personnel closer to the superintendent was in greater agreement than the more remote group; in 14 comparisons, the reverse occurred. This distribution is well within the bounds of chance. For the 38 comparisons of Consideration variances, 24 confirming instances were tabulated against 14 disconfirming instances, a distribution which also fails to exceed chance at the 5 per cent level.[21] Considering just the extremes of the contact groups, i.e., the administrative staff members and the category of teachers who, at best, only passed casual remarks with the superintendent every month or two, the former agreed less on the superintendent's Initiating Structure behavior than the most remote teachers. In the 13 school systems where the extreme comparisons could be made, score variances were higher among administrative staff members than among the remote teachers in 8 of them. With respect to Consideration behavior, the administrative staff agreed more in 9 of the 14 systems, less in only 5 of the 14. Neither of these last sets were statistically significant. In effect, those associates of the superintendent who had the least access to information about the superintendent's behavior nevertheless agreed on the character of that behavior as fully as the associates who could observe the behavior constantly. Such evidence does not support an interpretation that the LBDQ reports are veridical descriptions of the superintendent.

---

[21] The probabilities of finding at least 20 confirming instances in a sample of 34, and at least 24 in a sample of 38, where one-half of them would be expected on the basis of chance, is .20 and .07 respectively according to the exact test of the binomial distribution. Doubling these probability values (.40 and .14) provides the appropriate two-tailed test.

Summary

The preceding sections of this chapter have assembled evidence from the various studies of school superintendents bearing on the issue of whether or not LBDQ scores can be regarded as veridical descriptions of behavior. Below are the central findings.

1. No attempts have been made in the superintendent studies to relate LBDQ scores to theoretically relevant outside variables. Hence, it is impossible to report upon construct validity.
2. In some samples (e.g., teachers in the replication study), respondents were able to discriminate between two kinds of leader behavior in a given superintendent as represented by low correlations between Initiating Structure and Consideration scores. In other samples (notably administrative

staff members in the replication study), correlations between the dimensions were extremely high. Tests for internal consistency were not run in any of the superintendency studies.

3. Halo, as represented by substantial correlations between LBDQ scores and measures of generalized feelings about the superintendent, was found in the Consideration dimension but not the Initiating Structure dimension, a finding which corresponds to one reported by Seeman.
4. The Initiating Structure dimension includes several items which, if the respondent wishes to score the superintendent high, require a socially undesirable characterization of him. The Consideration dimension includes no such items, a fact which may serve to explain the difference between the two dimensions in their susceptibility to halo.
5. The "personal" form of the LBDQ, designed to encourage objective reporting on the part of teachers, was somewhat less imbued with halo than the "regular" form.
6. The nature of changes in teacher descriptions and superintendent self-descriptions over a three-year period are not consistent with an interpretation of veridicality. Teachers and superintendents disagreed on the direction in which the superintendents changed their behavior.
7. In each of the three superintendent studies, some categories of school personnel did not agree in their descriptions of

the same superintendent, and in one study active disagreement was recorded. In a few cases, substantial agreement between categories of respondents was found.

8. In the replication study, groups of personnel who had first-hand bases for knowing how their superintendent behaved were no more nor no less in agreement about his Initiating Structure or Consideration behavior than were personnel who had less access to information about him, including some who had virtually no contact at all with the superintendent.

The evidence in favor of the veridicality interpretation is far from compelling. It would appear that factors other than the objective fact of the superintendent's behavior determine the way in which school personnel complete the LBDQ forms. The next chapter will consider briefly an alternative interpretation of the LBDQ scores.

Chapter 8

Reinterpretation of the LBDQ Scores

If the LBDQ scores are not faithful representations of the behavior of superintendents, as the accumulated evidence seems to prove, what do they measure? While many teachers did not complete the LBDQ items (a point noted more fully in the next chapter), inspection of the returns indicated that they took the task seriously and, for the most part, responded with care. The scores cannot be attributed to random behavior on the part of respondents. The significant F ratios in the analyses of variance encourage us to reject the null hypothesis that the superintendents or those who describe the superintendents are random samples from the same population.

In this chapter we shall propose another position from which the LBDQ scores may be viewed--one which treats the objective fact of the superintendent's behavior as but one of two kinds of determinant which operate jointly to produce the LBDQ responses. Offered in this chapter, too, is an extremely limited analysis of data from the replication study designed to sort out the operation of the two kinds of determinant among teacher respondents.

Two Determinants of LBDQ Responses

Psychological theory suggests that the performance of respondents, as they attempt to describe the behavior of another person, is governed by two sets of forces. One set of forces is the objective stimulus situation--the facts of the behavior they seek to describe. The other set is within the respondents themselves--functional forces such as their needs, moods, feelings, and frustrations. The two sets are not independent of one another; both are involved in the rating performance of a respondent, but involved in varying degrees. When the objective stimulus situation is ambiguous and unclear, functional forces are allowed greater reign than when the objective facts are clear and compelling. When the functional forces within the respondent are strong, they may over-ride the facts from the outside world. His responses on a form such as the LBDQ may be more descriptive of him than of the person he attempts to describe.

On the grounds of psychological theory, then, the LBDQ scores produced by respondents are the products of both the facts about the superintendents' behavior and attributes of the respondents themselves. We have already seen in the preceding chapter that Consideration scores are concurrently predictable to an appreciable extent from teacher's global feelings about the superintendent. The correlation between Consideration scores and the graphic rating scale evaluation of superintendents' leadership quality was .59, with the measure of teacher affect toward the person of the superintendent, .55.

As we observed earlier, an alternative interpretation of these correlations is available. One could equally well say that teachers' feelings toward the superintendent are predictable from his Consideration behavior. Thus, it would have been the superintendent's behavior which created the affect in teachers, not the affect which shaped teachers' descriptions of his behavior.

A Test of the Joint Determinants

While any attempt to determine the direction of causality in ex-post-facto data is tenuous, we shall present some data which attempt to suggest the relative significance of objective facts and functional forces in LBDQ scores. The principal assumption involved in the analysis is that the objective stimulus situation--the superintendent's actual behavior--is more clear and compelling to persons in reasonably close contact with him than it is to those with little or no contact with him. If it is the objective facts of the behavior which most importantly govern respondents' descriptions of behavior, then the descriptions (the LBDQ scores, and Consideration scores in particular) should be less highly correlated with the affect measures among those in close contact than among those with little or no contact. That is to say, the clarity of the facts would tend to outweigh the affect level among those for whom the facts were the more clear. On the other hand, if it is the subjective level of affect which most importantly shapes teacher descriptions of the superintendent, the correlations between LBDQ scores and affect measures should be equally great among those with much contact and those with little or no contact with the superintendent.

In this tentative test, analyses were limited to data for teachers in the large school systems where substantial numbers had little contact with the superintendent. No data were available regarding affect levels among administrative staff members.

Furthermore, only P Scale scores (affect toward the superintendent) were used in the analysis. Teachers were dichotomized according to the procedure described in Chapter 6 to form "little contact" and "much contact" groups. Correlations between the P Scale scores and the LBDQ dimension scores ("regular" form) were computed separately in each group, with the results as shown in Table 26.

The Consideration score correlations are of particular interest in the table. Among teachers with "little contact" with the superintendent, affect toward the superintendent correlated .68 with Consideration scores; among those with "much contact," the correlation was .58. Both coefficients, based on N's of 55 and 56, are significantly different from zero beyond the 1 per cent level. But a test of the difference between the pair of coefficients showed them not to differ significantly. Analysis of this relationship, then, indicates that objective facts about the superintendents' behavior were not principally responsible for teachers' feelings toward them. Affect entered into teacher reports on Consideration as strongly among those with relatively great opportunities to furnish veridical descriptions as among those whose first-hand information about the superintendent was slight or non-existent.

The pair of correlations for the Initiating Structure dimension are interesting too. They are both smaller in magnitude than those for the Consideration dimension, a matter which our previous analyses would lead us to expect, but the difference between the pair again was found not to be statistically significant.

Table 26

Correlations of Teacher LBDQ Responses, "Regular" Form, with Teacher Affect toward the Superintendent, for Teachers with "Much" and "Little" Contact with their Superintendents: Large Systems Only

	LBDQ		
P Scale correlations among . . .	Initiating Structure	Consideration	Significance Levels Difference between Coefficients
"Much" contact teachers	.31*	.58**	$p > .05$
"Little" contact teachers	.18	.68**	$p > .05$

$^{*}p < .05$

$^{**}p < .01$

Note. The number of cases on which the correlations are based range from 55 to 58.

We must underscore the tentative nature of these analyses, based as they are on small numbers of cases--approximately 55 for each correlation--and uncontrolled as to other variables, such as length of service in the school system, which might account for the relationships. Moreover, our mode of reasoning leaves unresolved the question of the source of teachers' generalized feelings about the superintendent which colors their descriptions of him. If it is not first-hand observation of him which creates their affect toward him, what is it? The affect level could be a reflection of the teacher's momentary moods or more enduring personal frustrations and satisfactions, uncritically generalized to others in the working situation. There is another possible source, too. Certainly first-hand observation is not the only way the teacher has of "knowing" how the superintendent behaves. Important personages in the school system are subjects of conversations among teachers, and a considerable amount of information circulates about them. Social norms are bound to arise in different sectors of the faculty regarding what the superintendent is "really" like. This may account for the fact that many teachers who had never even seen the superintendent to talk with him were nevertheless willing to describe his behavior on the LBDQ. Data of the replication study, however, are not capable of illuminating the operation of social processes in the faculties as they influence teacher responses to the LBDQ.

In conclusion, we have suggested that any description of the superintendent is a compound of the objective facts of his behavior and subjective attributes of he who furnishes the description, and the scores must be interpreted in that way. In some instances, the objective facts may outweigh the functional attributes to produce relatively veridical reports, although we have not seen many such instances. In other cases, functional attributes of the respondents may weigh more heavily in the production of LBDQ scores. With regard to both organizational capacities and considerateness on the part of the superintendent, respondents' personal feelings appeared from our analyses to over-ride veridical descriptions of the superintendent's behavior. The reader should be warned, however, against placing strong credence in the results of the limited emperical analyses.

Chapter 9

Relevance of the LBDQ for the School Executive

As the investigator turned his attention to the item content of the LBDQ, as he was obliged to do when Hunter's findings failed to replicate, he began to question the relevance of the instrument for describing leadership in an organizational setting. Part of the questioning was directed specifically toward item wording in Hunter's revision of the LBDQ, but part was directed twoard the more general issue of the meaning of leadership in the bureaucratic organization. The present chapter will elaborate on these matters.

The original forms of the Leader Behavior Description Questionnaire were devised to measure leader behavior in small, face-to-face work groups. Items described kinds of behavior which a leader would undertake in direct, personal interaction with members of the group. To be sure, the LBDQ has been used to study leadership in organizational settings, but the focus of investigation has been on the leader's behavior with respect to his immediate staff, not the organization as a whole. In Halpin's study of the district superintendent, for example, it was the interaction between the superintendent and the seven school personnel with whom he worked most closely which the various parties were asked to describe.

Hunter's one-word revision of Halpin's items--"teacher" for "staff"--drastically altered the context of the leadership act and, we contend, made the descriptions largely meaningless. The district superintendent's role vis-a-vis the classroom teachers of the district is not that of a leader in a small task group, except possibly in very small districts (not represented in Hunter's study) where teachers report directly to the superintendent. We will expand on our contention in the first section to follow.

Beyond the matter of Hunter's revision is a more general question regarding the description of leadership in the setting of a bureaucratic organization. Leadership functions in the small work group are actively performed by members of the group and typically are centered in a single person identified as the leader. Leadership can be observed as someone's behavior, and the LBDQ is suited for its description. Many of the functions of leadership and administration in the bureaucratic organization, on the other hand, are institutionalized in the organizational structure and cannot be observed as someone's behavior at all. The functions are served by such impersonal mechanisms as written rules and regulations, clocks and bells which initiate activity, and the established division of labor which allocates tasks. Leadership functions servedin these ways cannot be regarded as behavior except by a severe stretch of meaning and, in any event, cannot be attributed to a single person. To the extent that leadership is built into the structure of the organization, a description of leader behavior is irrelevant. At best it describes residual aspects of leadership. This limitation of the LBDQ, which applies with equal force to the first line supervisor, the principal, and to the executive office of the superintendent, will be considered in greater detail in the second section.

Role of the Superintendent vis-a-vis Teachers

That the district superintendent's relationship with teachers is wholly unlike the relationship between leader and members of a small, face-to-face work group is so patent an observation that an extended discussion would belabor the obvious. We will comment on several aspects of the relationship not simply to justify the observation but to illuminate the nature of the superintendent-teacher relationship in the school systems under investigation. First, we turn to the fact that in these systems, where at least one level of authority intervenes between the teacher and the superintendent, the superintendent's functional connection with teachers is extremely limited. Next we examine the contact data from teachers bearing upon their remoteness from the district superintendent. Especially illuminating are the break-downs of these data for school systems varying in size, discussed in the succeeding section. Finally, we return to the LBDQ responses, or better the non-responses, to suggest the difficulty which teachers had in describing the superintendent's behavior by means of the LBDQ.

Intervening levels of authority. In a school system as large as the smallest in Hunter's study or in the replication study, the superintendent's over-arching responsibility for instructional operations normally is delegated to an intermediate officer--the building principal. The removal of the superintendent from day to day supervisory relations with classroom teachers was documented in the superintendent interviews. When asked how much time he spends with the teachers, the superintendent of one of the smaller school systems said: "Well, I try to eat lunch with them every day, but I don't spend as much time as I used to. I don't want to undermine the principal."

The smallest school system of the 20 in the replication study, with 33 classroom teachers, had recently changed from a one-man administrative structure to a three-man administration upon the appointment of building principals in the high school and the elementary school. The superintendent reported:

> I spend much less time now with teachers than it used to be. For one thing, I am attending a lot more meetings outside in the district. So more and more I've turned the details over to the principals. If I want to see a teacher, I arrange to spend some time in the teachers' lounge until I can catch him. I attend none of the faculty meetings. The main thing is that I do that to increase the position of the principals.

In another of the smallest systems the superintendent had this to say about the time he spends with teachers:

> Very little time; perhaps two or three hours a week at the most. There are problems and interruptions coming up all the time. Right now I have all the headaches of the new high school building. It seems like I draw more and more away from teaching, from supervision.

In the larger systems, of course, the superintendent's tasks remove him even further from classroom teachers. Where the system is large enough to support a central office staff, the superintendent' functions may be restricted to the coordination of the staff specialists who are responsible for the various administrative duties and they may spend more of this time in external relations of the school system.

Running through the interview protocols in the replication study were three principal occasions of which superintendents' tasks brought them into contact with teachers. Most of the superintendents interviewed teachers when they were hired in the system and talked with them at the time of their resignation. Apparently there were occasions in the intervening period when teachers would come to the superintendent concerning their professional plans.

A second circumstance which brought teachers and superintendent together was the negotiation of salary adjustments. In systems where the superintendent mentioned this, the superintendent met with a group of teacher delegates to work out salary recommendations to the board of education.

A third occasion for contact, mentioned especially in the larger systems, was the superintendent's participation incurriculum committees. While our evidence is not complete, it seems that the participation was more in the form of providing symbolic support for the curriculum effort than of giving active leadership.

The remarks of the superintendent of a large and rapidly growing suburban school district gives substance to the above three points, although his withdrawal from the employment area was not typical:

> I've looked in on some of the curriculum studies. But about half of the teachers nowadays don't even see the superintendents in hiring. It's different for the older teachers who were hired by the superintendent. I know some of them, and I see them once in a while. Then there is the salary committee once or twice in the fall. But practically one hundred per cent of the job is administrative . . . I don't even see much of the principals. All the work is done through administrative assistants. That's their job. I never see teachers.

It is abundantly clear that the relationship between superintendent and teacher is not akin to that of a leader in a face-to-face task group upon which the LBDQ is predicated, not even in the smallest school systems of the application study. Although our evidence is not sufficient to test it, the investigator was impressed by the possibility that small system superinendents, lacking administrative aids, were more subject to work demands which removed them from interaction with teachers than superintendents in the larger systems. But in any case, it was the building principals, not the superintendent, upon whom supervisory responsibilities rested.

Remoteness of teachers from the superintendent. It is, of course, a physical impossibility for a superintendent to interact regularly with even a majority of his teaching faculty through the school year in any reasonably large school system. Data from teacher questionnaires regarding their contacts with the superintendent provide a direct means for estimating how the superintendents distributed their contacts with the faculties. Fully 70 per cent of all reporting teachers said they saw their superintendents to talk with them no more frequently than every month or so; indeed

30 per cent said they saw him not more than twice a year if they saw him at all. These percentage figures are shown in Table 27 and are based upon the frequency figures in Table 17, Chapter 6. Some 40 per cent of all teachers indicated that if they had any occasion to talk with the superintendent, it was limited to "brief, casual remarks." Less than one-quarter of the teachers fell in what we have called the High Intimacy-High Frequency category. If leadership requires personal interaction, few of the teachers in the systems under investigation could regard the superintendent as their leader.

Remoteness and school system size. The raw size of a school system--the number of teachers on the faculty--certainly would have an effect on the superintendent's opportunity to distribute his contacts widely across the faculty. Table 28 summarizes the contact distributions separately for large and small school systems. Only 5 per cent of the teachers in the large school systems were in the highest contact category (High Intimacy-High Frequency), whereas 41 per cent of teachers in the small systems fell in this category.

The weighted contact scores for each school system, also shown in Table 28, display an interesting uniformity in superintendent-teacher contacts in the large systems which is absent in the small systems. Contact scores range only from .26 to .81 in the large systems--around a mean of .51 for the size class--while the contact scores for the small systems are spread from .68 to 1.87. The maximum attainable score is 2.00. One of the small systems (Code

Table 27

Percentage of Teachers with Varying Degrees of Intimacy and Frequency of Contact with the Superintendent

	Frequency of Contact				
Intimacy of Contact	Less than once, twice yearly or none	Every several months	Several times monthly or weekly	Total	(Number of Teachers)
Brief remarks or no contact	21%	14%	6%	41%	(263)
More intimate contact	10%	26%	23%	59%	(386)
Total	31%	40%	29%	100%	
(Number of Teachers)	(199)	(262)	(188)		(649)*

---

*Excluding 13 of 662 teachers whose responses either were inconsistent or could not be ascertained.

Note. The more detailed frequencies are shown in Table 17 of Chapter 6.

Table 28

Teacher Contact with the Superintendent: Percentage Distributions over Contact Index for Large and Small School Systems and Weighted Scores for Each of Twenty Systems

Contact Index		Per Cent	
Intimacy	Frequency	Large	Small
High	High	5%	41%
High	Low	40	33
Low	Low	53	16
Low	High	2	10
Total		100%	100%
(Number of Cases)		(325)	(324)

Large School Code	Contact Score	Small School Code	Contact Score
00	.26	10	1.63
01	.42	11	1.30
02	.81	12	.87
03	.50	13	1.74
04	.52	14	1.87
05	.45	15	1.31
06	.58	16	.68
07	.46	17	1.00
08	.53	18	1.55
09	.55	19	.93
Size Class	.51	Size Class	1.28

16) looks more like a large system in its pattern of teacher contacts than like the other small systems. In another small system (Code 14), four-fifths of the teachers were in the highest contact category, giving it a score of 1.87.

From another perspective, however, the difference in superintendent-teacher contacts between large and small systems is not so striking. It is possible to estimate the raw numbers of teachers in each school system with whom the superintendent has frequent and intimate contact by multiplying the total number of teachers in the school system (not just the number in the sample) by the percentage of the sample in the High Intimacy-High Frequency category.[1] These estimates, shown in Table 29, indicate only slight differences between the large and small system averages--15 and 19 teachers respectively. While the superintendent who interacted regularly with the greatest number of teachers--46 of them--was in one of the small systems, those with the second and third highest numbers were in large systems, the latter in the largest system of all.

---

[1] The estimates are subject to two kinds of sampling error, varying from one system to the next. The first is the sampling rate. In large school systems the sample of 40 teachers represented a small proportion of the total faculty--approximately one in ten teachers in the largest system. In the small systems, well over one-half of the teachers were sampled, and in five of them the total staff was taken as the sample. The second source of error is the return rate of teacher questionnaires. Of the teachers sampled, roughly one in ten did not return usable questionnaires. Again, the rate of returns varied from one system to another, so that in one system the rate was 75 per cent and in another it was 100 per cent. In the latter case, the entire faculty was sampled and, as a result, the estimate of the number of teachers with frequent and intimate interaction with the superintendent was subject to neither kind of sampling error.

Table 29

Estimated Raw Number of Teachers in Each School System with Whom Superintendent had Intimate, Frequent Contacts

School System Code	Per Cent of Sample in Quadrant 3*	Total Number of Teachers in System	Estimated Number of Teachers Contacted by Superintendent
Large Systems			
00	6%	275	16
01	3	294	9
02	14	243	34
03	6	306	18
04	-	225	0
05	3	384	12
06	4	362	14
07	6	425	25
08	3	358	11
09	3	354	11
Mean			15.0
Small Systems			
10	61%	38	25
11	33	35	12
12	21	61	13
13	76	60	46
14	81	33	27
15	43	39	17
16	3	49	1
17	20	62	12
18	50	41	20
19	27	63	17
Mean			19.0

---

*Quadrant 3 corresponds to the High Intimacy-High Frequency category.

Other evidence, too, suggests that the difference between large and small systems in the distributions of superintendents' contacts with their faculties was a product of accidental features of the settings of the systems rather than of a difference in the superintendent's role vis-a-vis teachers.

In every one of the large systems, for example, the superintendent's office was situated in a separate administration building, sometimes near the high school but more often in a location geographically remote from any school building. His office in the small systems was usually located in one of the school buildings or in the school building. Only one of the small systems (Code 16) housed its superintendent completely apart from any building, and it was this system in which the rate of teacher contact was most akin to rates in the large systems. To examine the effect of geographical location more closely, teachers in the small systems were categorized according to whether or not they taught in the same building as that in which the superintendent was housed. About one-half of them did. Their contact scores were substantially higher than those of teachers located apart from the superintendent--scores of 1.49 as opposed to 1.06.

Our data also show that more teachers in the small school systems were acquainted with the superintendent on a personal level than in the large systems. All but three of the small systems (Code 12, 16, and 19) were located in small towns or rural neighborhoods where opportunites for knowing the superintendent and his family on grounds other than professional connections were great. The three exceptions, small districts in highly urbanized areas of St. Louis County, rank the lowest in teacher contacts of all the small systems.

The effect of personal acquaintance on contact with the superintendent was examined in another, more detailed way. Two questions on the teacher questionnaire asked teachers about their personal acquaintance with top school officials: "Have any of them visited in your home?" and "Are you on a first-name basis with any of them?" While these questions did not ask directly about personal acquaintance with the school superintendent, they furnish a foundation for constructing a general index of particularistic relationships between teachers and school officers which cuts across the hierarchical order of the work situation. Teachers were coded 2 if they answered "yes" to both questions, 1 if they answered "no" to the first but "yes" to the second, and 0 if they answered "no" to both. (Only a handful of teachers answered "yes" to the first and "no" to the second and were ignored in our analyses.)

Where 34 per cent of the teachers in the small systems were personally acquainted with top school officials (had been visited by them and were on a first-name basis with them), only 11 per cent of the large system teachers could claim this degree of acquaintance with them. (See Table 30.) Not surprisingly, the closer the teacher's personal acquaintance with the superintendent, the higher

Table 30

Percentage and Mean Contact Scores of Teachers with Varying Degrees of Personal Acquaintance with the Superintendent, in Large and Small School Systems

Personal Acquaintance		Per Cent		Mean Contact Score	
Visited	First-name	Large	Small	Large	Small
yes	yes	11%	34%	.87	1.52
no	yes	16	22	.66	1.31
no	no	73	44	.42	1.11
Mean				.50	1.29
Total		100%	100%		
(Number of teachers)*		(302)	(274)		

---

*Figures exclude 86 of 662 teachers for whom Contact Scores could not be calculated and degree of personal acquaintance was not ascertained.

was his rate of contact with the superintendent, as the contact scores in the last two columns of Table 30 reveal. This was true for both the small and the large systems. The fact that the teacher-superintendent interaction remained consistently higher in the small than in the large systems at each level of acquaintance suggests that something else militates against contact in the large school systems, probably the factor of geographical separation noted above.

In sum, the evidence points to two factors which contribute to the difference between the rate of superintendent-teacher contacts in the large and small school systems. The difference is in part a function of the proximity of the superintendent's office to the work location of the teachers and in part a function of the opportunities offered by small town settings of the school district for interaction unrelated to the work situation. The fact remains that only in two school systems did more than two-thirds of the teaching faculty have any regular, significant interaction with the superintendent, whatever the basis for it might have been. The data on teacher contacts, alone, make it obvious that the superintendent-teacher dyad does not correspond to a leader-follower relationship in the sense of a small, face-to-face task group. The difficulties which teachers experienced in attempting to respond to the LBDQ--whose items imply that such correspondence exists--constitutes a final source of evidence on the point.

Difficulties in responding to the LBDQ. Many of the teachers replying to the Leader Behavior Description Questionnaire with regard to their superintendents apparently found the items difficult to answer or even meaningless for them. One-quarter of the 662 respondents either omitted items or skipped the entire LBDQ schedule so that no score could be derived for them on either the Initiating Structure or the Consideration dimension. No other part of the questionnaire, including the 30 personally probing items on the back of the form regarding their feelings about their work situation, had such a high proportion of incomplete responses.

Moreover, the proportion of incomplete responses was higher on the "personal" form of the LBDQ, which required specific reports of how the superintendents related to the responding teachers individually, than on the vaguely worded "regular" form. Unscorable Initiating Structure and Consideration responses for the "personal" form were 27 per cent and 29 per cent respectively. They were 24 per cent and 21 per cent respectively on the "regular" form which asked for descriptions of the superintendent's relations with "teachers in general."

Unsolicited comments written on the questionnaires were more abundant on the LBDQ schedule than on any other part. A total of 125 returns contained comments regarding these items, mainly in the vein of "don't know," "can't answer," and"haven't seen the superintendent enough to make a judgment." A number of respondents expressly stated that the questions applied not to their superintendent but to their building principal or some other lower echelon administrative officer. The following comments are illustrative.

> I'm a bit confused as to how to answer these questions mainly because of the fact that our school principal is the one who I have contact with. Our superintendent couldn't possibly do many of these things for me because he's got 495 teachers and 16 schools to oversee.

Another large system teacher wrote,

> We have two assistant superintendents and many special supervisors that take care of about all the above listed items so that it does not require direct contact with him as it would in a smaller system.

But even in the small systems the teachers pointed out the inappropriateness of the items when applied to the superintendent.

> These statements hardly apply directly to our superintendent as faculty procedure, meetings etc. are called and controlled by principal.
>
> Our superintendent is friendly and approachable, but we do not go over our principal's head with anything--little contact with superintendent.
>
> . . . We are responsible to the principal and assistant principal.

Comments such as the above were somewhat less common on questionnaires from teachers in the small systems. Further, the number of unscorable LBDQ schedules were fewer in the small systems--20 per cent compared with 30 per cent in the large systems--reflecting, possibly, a closer relationship between teacher and superintendent. But the sum of the evidence presented in this and previous sections indicates the general irrelevance of the LBDQ items for describing the superintendent's leadership.

The Institutionalization of Leadership

The foregoing observations regarding the superintendent's role *vis-a-vis* teachers underscore the "lack of fit" of Hunter's revised LBDQ items for describing the superintendent's leader behavior. They suggest the unclarity which must have been in the teachers' minds as they attempted to answer the questions, which, in turn, helps to explain why the responses were found to be so heavily endowed with halo. An issue of broader consequence than the problems associated with the LBDQ, however, demands some comment. To what extent is the concept of "leadership" or "leader behavior" appropriate at all in a bureaucratic setting such as the school?

The referent of "leadership" is highly personalistic. The concept points to actions of individual human beings which affect the behavior of others as they jointly pursue a task. And yet it seems that relatively little of the work carried out by school personnel is the result of direct instigation by the acts of a particular "leader." Except possibly in the case of *ad hoc* work groups assembled for special, limited purposes, performance is governed more by the demands of an impersonal work structure than by the directives of a principal, a superintendent, or anyone else. It is exactly this established work structure which enables the organization to persist through time in the face of replacement of personnel and even in the absence of administrative officers.

The structure of organization. Many of the functions personally served by the behavior of a leader in the small group are, in the complex bureaucracy, built into the structure of the organization.

This can best be illustrated in reference to the dimensions and items of the LBDQ.

Consider the definition of Initiating Structure, referring to "the leader's behavior in delineating the relationship between himself and members of the work-group, and in endeavoring to establish well-defined patterns of organization, channels of communication, and methods of procedure."[2] To a large extent these matters are already delineated and established in an enduring organization. The school officer, for example, need not delineate his relationships with others; the essential aspects of his role are specified by his position in the school's authority structure. He may clarify or even alter the established role relationships, but only within the limits of the gross structure. It is the rare occasion when any given officer completely revises the authority structure of the school system. Short of that, the personal efforts of an officer in delineating his relationships are residual contributions, not primary ones.

No particular school "leader" establishes _de novo_ the patterns of school organization. The fundamental outlines of the work organization are institutionalized in the educational division of labor--are part and parcel of patterns of teacher training, teacher certification laws, and broadly shared understandings of the teachers's role. The functions envisaged by the following LBDQ items (Hunter revision) may be appropriate in the ad hoc work group but hardly in the school system:

---

[2] Andrew W. Halpin, _Leadership Behavior of School Superintendents_, School Community Development Monographs No. 4, Columbus, Ohio: Ohio State University Press, 1957. P. 4.

Assigns teachers to particular tasks

Lets teachers know what is expected of them.

It is the job title and employment contract which tell the third grade teacher that he is to teach third grade and the high school science teacher that he is to teach high school science, and both know in general outline what the tasks involve. The knowledge is available in the culture of society and affirmed and further specified in teacher training. Details of the teacher's task suited to a particular school system are likely to be furnished by impersonal means--the curriculum guide or the textbook series. While re-organization of the work structure occasionally occurs, and can be attributed to the personal efforts of a principal or superintendent, it must always be within the limits of the institutional framework, and once the re-organization has been accomplished, the personal contributions of the "leader" recede and the contributions of structure replace them.

Methods of procedure are normal aspects of complex organization. They are codified in the teacher's handbook, in accident reporting requirements, in attendance forms, in the grading system, and the like, and it would be difficult for a teacher or anyone else to attribute their presence to a given administrative officer. Standard operating procedures and written rules and regulations are bureaucratic substitutes for such acts of work group leaders as:

Emphasizes the meeting of deadlines for teachers

Ask that teachers follow standard rules and regulations.

Encourages the use of uniform procedures.

An executive may be either lenient or strict in their enforcement, thereby making a difference which can be explicitly attributed to him, but the lattitude open to him is relatively minor in comparison with the massive effects upon teachers of the impersonal, commonly accepted standard operating procedures themselves. A building principal, for example, may insist that teachers arrive at school 15 minutes before classes take up, but the more powerful fact--a wholly impersonal one--is that instruction begins in all schools of the system at 8:30 a.m. Teachers, bus drivers, parents, pupils, and crossing patrolmen are all geared to this massive fact and would respond even if the building principal took a year's leave of absence.

Professional norms. Another form of behavior envisaged for the work group leader is that of inducing a high quality of performance among the members. Relevant LBDQ items are:

Sees to it that teachers are working to capacity.

Maintains definite standards of performance for teachers.

Critizes poor work by teachers.

The mechanisms for accomplishing this function in the school system again are impersonal ones and, for the most part, rest not in the organizational structure at all but in the norms of the occupational group and in the teacher who internalizes them. Much of the teacher's performance is inaccessible to observation, and he is commonly afforded a high degree of autonomy in his modes of work. Under these circumstances, the motivation to perform must be sought in the teacher himself, not in supervisory practices.

The teacher's motivation is reinforced by a professional or service-oriented ideology in which his colleagues share, and it is they rather than school officers who are in a position to exercise social control over the teacher. The questions cited above regarding the superintendent's supervisory behavior excited a number of written notations on the LBDQ forms. "He knows I always work up to capacity," "Only if you mean constructive criticism," and "Certainly not!!!" are illustrative. Apparently respondents saw the implication underlying these items as violating the norms concerning their status as professional workers.

The school system's involvement in inducing high quality performance seems to be minimal. Teachers are presumed at the time of their employment to be competent to the extent of their prior experience and training. The rewards for competence, codified in salary schedules, are distributed largely on these grounds. While there are exceptions to the foregoing observations, it is nevertheless obvious that many of the functions of inducing teacher performance are outside the jurisdiction of any given officer, including the school superintendent.

Consideration in bureaucratic organization. It is easy to see the way in which the functions of leadership associated with the Initiating Structure dimension of the LBDQ are imbedded in impersonal mechanisms of bureaucratic organization. Less amenable to institutionalization, however, are the "friendship, mutual trust, respect, and warmth in the relationship" aspects of the Consideration dimension. Certainly the "leader" in the schools can contribute substantially to the organizational climate in which the work is carried out.[3] And yet the bureaucratic setting casts some of the elements of the leader-follower relationship in a different light from that of the small work group.

Friendship, for example, carries certain dangers in the operation of large organizations which may not be so problemmatic in the intimate work group. Three items of the LBDQ describing the superintendent's behavior read:

> Looks out for the personal welfare of individual teachers.
> Does little things to make it pleasant to be a teacher.
> Does personal favors for teachers.

In the setting of large-scale organization, such behaviors on the part of a ranking officer may take on the character of paternalism or worse, favoritism. According to the written comments, a number of teacher respondents took the latter item to mean exactly that--favoritism. From their perspective it was a highly improper mode of behavior. Any particularistic relationship between the school officer and his faculty violates the universalistic norms by which the bureaucratic organization gains its strength. "Warm, friendly relationships" must exist in a highly qualified form where an authority structure is interposed between leader and follower.

---

[3] Andrew W. Halpin, The Organizational Climate of Schools, Final report, Cooperative Research Project No. 543, U. S. Office of Education, 1962.

Some of the items keyed for the Consideration dimension of the LBDQ relate to teacher involvement in the school system's decision processes.

> Acts without consulting his teachers. (Negative)
>
> Gets teacher approval on important matters before proceeding.
>
> Refuses to explain his actions to teachers. (Negative)

Decision functions are more highly specialized in the bureaucratic organization than in the small work group. Both teachers and administrators recognize that there are realms in which it is the administrator's responsibility to make decisions without obtaining "clearance" from the teaching faculty. In fact, his unwillingness to make decisions independently is often regarded as his avoiding responsibility.

There are other realms in which decision-making without consulting the affected teachers would be regarded as improper. In the LBDQ forms, the item regarding "getting teacher approval" was frequently qualified by respondents with the comment, "when they are important to the teachers." But even here teacher participation in administrative decision making may be a matter of organizational structure, not leader option. Some respondents pointed out the existence of welfare committees or administrative councils which formally involved teachers in the decision-making process. Where such organizational arrangements are well established, it would be difficult to attribute teacher participation to the behavior of a given school officer, unless he happened to be the person primarily responsible for the structural arrangement in the first place.

That bureaucratic organization may approach the problem of considerateness for the individual with the same rationality as it invokes in its task-oriented functions is the principal implication which emerged from Moeller's companion investigation.

> . . . Every practical procedure will be followed by the bureaucratic school system to enable teachers to do their best work. Special efforts to make teachers feel a part of the system, including orientation meetings stressing fellowship, tours of the community and even finding new teachers a place in which to live, are some of the procedures used to integrate new teachers into the system and to keep their state of emotional well-being at such a point that the teachers will want to remain in the system.[4]

Rationality means the embodiment of these functions in stable, structural arrangements, free of the preferences and whims of particular persons who may inhabit the administrative offices. While friendship and warmth cannot be assured by organizational structure, the structural pattern can communicate to the participant the extent to which members of his category are trusted and respected for their contributions. A different organizational form appears when it is constructed on the premise of the worth and dignity of each participant than when it is shaped by an alternative premise. To attribute the form of the organization to one palpable official such as a superintendent or principal, however, is exceedingly difficult. It may be possible in the exceptional case in which the official has truly initiated the structure, but this is exceptional.

---

[4] Gerald H. Moeller, The Relationship between Bureaucracy in School System Organization and Teachers' Sense of Power, Unpublished doctoral dissertation, Washington University, 1962. Pp. 141-142.

To repeat, we believe that leadership is not the same in the large, complex organization as in the small *ad hoc* work group. Where the functions of leadership can only be accomplished in the latter through the behavioral acts of individuals, they are often accomplished in the bureaucratic setting by totally impersonal mechanisms and procedures. In one conceptualization of the contributions of superintendents to school system effectiveness, Halpin differentiated between two kinds of behavior: behavior as a leader *vis-a-vis* the superintendent's immediate administrative staff, and behavior as a problem solver or decision maker *vis-a-vis* the teaching faculty and the school system at large.[5] He would restrict descriptions of the superintendent's Initiating Structure and Consideration behavior to the superintendent-staff aspect of analysis.

But even this formulation places what may be an undue emphasis upon the personalistic as opposed to the structuralistic features of the school system. How much of the school's effectiveness is contributed by the person of the superintendent and his varying capacities in decision making? How much is contributed by the varying manifestations of the work structure? This is a question

---

[4] Andrew W. Halpin, "A Paradigm for Research on Administrator Behavior," in Roald F. Campbell and Russell T. Gregg (Eds.), *Administrative Behavior in Education*, New York: Harper, 1957. Esp. pp. 166-173.

which cannot be settled at the present time, but the investigator believes that the significance of the structuralistic component all too often is overlooked.

Summary

We have developed the theme in this chapter that the superintendent-teacher relationship is totally unlike the leader-follower relationship whose dimensions the LBDQ was designed originally to describe. Interviews with superintendents documented the lack of close functional connectedness between their roles and the role of teacher, at least in school systems sufficiently large to support an intermediate authority level, as was true in all of the systems in the replication study. The roles of teacher and superintendent were found to connect on three occasions--on the teacher's employment and resignation, in salary negotiations, and when the superintendent participated in curriculum committees. In all three circumstances, the relationship seemed to be fleeting and, except for the first, limited to certain members of the faculty. It was the principal or some other intermediate official whose roles interlocked most closely with teachers in large and small systems alike.

Data from teachers regarding their contacts with the superintendent indicated that superintendents interacted regularly with less than a score of the classroom teachers in the system, a number which represents a fairly high proportion of the total faculty in a small system but only a small fraction of the faculty in a large system. The distribution of the superintendent's contacts across his faculty was shown to be affected by two factors, the geographical location of the superintendent's office with respect to school buildings and the urban versus small town setting of the system as this affected the opportunity for teachers and superintendents to become acquainted on a non-professional basis. Perhaps the most interesting feature of these data, one which we did not pursue, was the marked variation from one superintendent to the next in extent of their contacts with teachers.

Inspection of teacher responses to the LBDQ schedule--particularly their unsolicited comments and the number of incomplete returns--pointed to the general inapplicability of the LBDQ items for describing superintendent behavior toward teachers. Twenty to thirty per cent of the schedules were so incomplete that no scores could be computed. The higher percentage was for the "personal" form which required teachers to indicate the mode of the superintendent's behavior toward them individually; the lower percentage for the more vaguely worded "regular" form. Teachers' voluntary comments confirmed our earlier observation: the items could apply to the principal, but to the superintendent, no. Considering the patent absurdity of some of the items,[6] the remarkable fact is the number of teachers who did complete the schedules. It would seem that sufficient information about the school's executive officer circulates through the informal communication networks to afford even the more remote teachers a basis for making confident statements about him.

---

[6] Especially so in the "personal" form. Conceive of an inexperienced teacher in a system of some 300 teachers, for example, describing the frequency with which the superintendent "gets my approval on important matters before proceeding."

Although the relationship between the building principal and the classroom teacher is a closer approximation to a leader-follower relation, even here the question must be raised, and was raised in the latter part of the chapter, as to how many of the leader's functions are served by behavioral acts of an official and how many are embedded in the very structure of the organization. Reference to the LBDQ items and dimension definitions indicated in a variety of ways the manner in which impersonal mechanisms of organization and the norms of the profession serve the same purposes which, in the small work group, are met by the activities of the leader. This is most evident in the area of the LBDQ covered by the Initiating Structure dimension. Consideration for the well-being of followers would seem to be more difficult to institutionalize. But the argument was developed that certain aspects of Consideration may, indeed, be embedded in formal procedures of the organization and that certain other aspects, especially "warm, friendly relations" between superior and subordinate, must take on a different meaning in the bureaucratic context or else threaten to violate the universalistic norms by which the organization operates.

In short, the LBDQ as revised and used by Hunter and re-used in the replication study is not relevant for describing the school superintendent, a fact which helps to explain the invalidity of the instrument as documented in Chapter 7. Moreover, any procedure which seeks to describe the behavior of a school executive without taking into account the massive contribution to "leadership" made by the organizational structure itself may end by revealing only the residual and trivial features of the governance of cooperative endeavor.

Chapter 10

Summary of Parts I and II

While investigating a problem in interpersonal perception in 1958, Hunter[1] encountered an unanticipated pattern of findings concerning the way school personnel described the leadership behavior of the district superintendent, a pattern which was associated with school system size. Teachers and board members in the large school systems of his study (6,500 to 10,200 pupils) described their superintendents lower on Initiating Structure and Consideration, the two dimensions of the Leader Behavior Description Questionnaire (LBDQ), than did their counterparts in the relatively small systems (750 to 2,700 pupils). At the same time, the self-descriptions of superintendents in the large systems were markedly higher on the two LBDQ dimensions than those of superintendents in the small systems. All comparisons between respondents in large and small systems were statistically significant, except the difference for school board members on the Consideration dimension. The larger the school system, the lower were the teacher and board descriptions but the higher were the superintendent self-descriptions.

---

[1] Ovid N. Hunter, Relationship Between School Size and Discrepancy in Perception of the Superintendent's Behavior, Unpublished doctoral dissertation, Washington University, 1959.

Had all three sets of respondent differed in the same direction, the findings would not have been so interesting. One might readily have interpreted them as indicating a true difference in superintendents' leader behavior associated with the size of school systems. It was the contradictory character of the responses which made the findings unusual. To explain them, one would have to abandon the assumption that the LBDQ scores directly reflect how the superintendents in fact behave and entertain an alternative assumption that the scores unintentionally reflected some regularity in the psychological processes of the respondents which, in turn, varied with characteristics of the school organization. Contrary as the alternative assumption is to the premise of the LBDQ, its examination offered the prospect of learning about the psychological consequences of organizational structure.

The Preliminary Search for Confirmation

Hunter's dissertation study, devoted as it was to an entirely different problem from the examination of such an alternative assumption, furnished no substantial data which could throw further light on the question. Consequently, the present writer pursued the matter in several ways, initially to discover whether other studies in school settings yielded similar results and eventually by conducting a full-fledged replication of Hunter's investigation.

First, he consulted the LBDQ scores reported by Andrew W. Halpin in his monograph on 50 Ohio school systems and their superintendents in conjunction with size data supplied by Professor Halpin.[2] In several ways the Hunter and the Halpin data were not comparable. For example, where Hunter had obtained responses from classroom teachers, Halpin had used persons whom the superintendents reported to be their closest subordinate co-workers, whether teachers, secretaries, or administrative officers. Also, there was not close comparability in the size ranges of the school systems covered in the two studies, especially at the upper size limits. Nevertheless it was possible to match 14 of the Ohio systems with Hunter's "small" category and roughly equate 5 of them with Hunter's "large" category. Inspection of the Initiating Structure and Consideration scores furnished by Halpin's various respondents revealed no trace of the pattern of differences discovered by Hunter nor, indeed, any systematic differences associated with school system size.

Professor Halpin also furnished the writer with some analysis he had made on other data regarding the principals' leadership behavior.[3] These data had little comparability with those of Hunter, but, again, there was no indication of variations in LBDQ scores with school size.

---

[2] Andrew W. Halpin, Leadership Behavior of School Superintendents, School Community Development Monograph No. 4, Columbus, Ohio: Ohio State University Press, 1957.

[3] Warren Evenson, The Leadership Behavior of High School Principals: Perceptions and Expectations of Superintendents, Principals and Staff Members, Unpublished doctoral dissertation, University of Chicago, 1958.

There was an additional point of disparity between Hunter's investigation and Halpin's Ohio study, one which seemed trivial at first glance. Although Hunter had patterned his LBDQ items after Halpin (who, in turn, had adapted them from the measure developed in the Personnel Research Board investigations at Ohio State University[4]), Hunter had systematically changed one word in each item to adapt it to his purpose. He replaced the word "staff" with "teacher." For example, where one of Halpin's items read, "The superintendent assigns staff members to particular tasks," Hunter altered it to read, ". . . assigns teachers to particular tasks." The change fit Hunter's plans since he intended to use classroom teachers as his main group of respondents and did not plan to collect data from groups corresponding to Halpin's "staff"--i.e., the superintendents' closest co-workers. But the change also drastically altered the significance of certain of the items. No longer did they describe a kind of behavior which normally is in the superintendent's province of responsibility, especially, we suspected, in the larger of the school systems. Some of the items, such as "assigning teachers to particular tasks," seemed to describe responsibilities of the building principal rather than the superintendent in school systems with several hundred teachers, although in small school systems the superintendent might well undertake them. Other items were less affected by the change and did not seem to harbor the implicit "bias" against the large system superintendent.

---

[4] Ralph M. Stogdill and Alvin E. Coons, (Eds.) Leader Behavior: Its Description and Measurement, Research Monograph No. 88, Columbus, Ohio: Bureau of Business Research, Ohio State University Press, 1957.

The writer re-analyzed Hunter's data from teacher respondents with the thought in mind that differences could be attributed to the "bias" of the revised items against the large system superintendent. Items were classified on _a priori_ criteria regarding the extent of "bias" which they incorporated, and teacher scores were computed over the "biased" and "unbiased" items, then compared by school system size. The results disconfirmed the writer's hypothesis. The same size difference which Hunter originally had noted appeared in scores on the "unbiased" as well as the "biased" items. Moreover, an item-by-item inspection of response distributions, regardless of _a priori_ criteria, revealed no particular sub-set of items to which the size difference could be attributed; it existed in virtually all of them. Evidence collected later in the replication study suggested why this otherwise reasonable hypothesis failed: the superintendent's relationship to teachers in the so-called small systems was highly similar to that in the larger systems. In effect, the items were equally "biased" against the superintendents in large and small school systems.

The Replication of Hunter's Investigation

At this juncture, the writer was awarded a contract with the Cooperative Research Program of the U. S. Office of Education to replicate the Hunter study as closely as possible and, if Hunter's pattern of findings were substantiated, to attempt to account for them. LBDQ scores were obtained from classroom teachers, from administrative staff members (to approximate Halpin's category of "staff"), and from superintendents themselves in 20 school systems of about the same size and in the same geographical area(the St. Louis metropolitan region) as in Hunter's study. Data were not taken, however, from school board members. Eight of the school systems were identical to the ones which Hunter had studied three years previously and were still governed by the same superintendents. Questionnaires and self-return envelopes were mailed to approximately 40 teachers in each school system, selected randomly from school rosters, and to administrative staff members nominated by superintendents as their closest subordinate co-workers. In the case of superintendents, LBDQ forms were mailed to them but picked up by the research staff at the time personal interviews were held with them. Replies were received from 88 per cent of the 783 sampled teachers and 96 per cent of the 109 administrative staff members. Nineteen of the 20 superintendents submitted scorable LBDQ schedules.

Once the responses were tabulated, it became clear that the remarkable pattern of findings which Hunter had observed in his 1958 investigation had disappeared in the ensuing three years. They were not substantiated. Of the four comparisons which could be made between large and small systems--Initiating Structure and Consideration scores from teachers and from superintendents--only one was statistically significant and that in the opposite direction from what Hunter had found. Teachers in the large systems now described their superintendents as higher on the Initiating Structure dimension that those in the small systems.

Interestingly, scores from administrative staff members in the large school systems also were significantly higher than in the small systems on both dimensions of the LBDQ, although there was no group of respondents in the Hunter investigation against which this finding could be compared.

Since we had data from some of the same systems regarding the same superintendents as Hunter had studied, a direct comparison of responses could be made across the three-year period. Not only did the size differences fail to appear in the 1961 data, although they had been present for this sub-sample of systems in 1958, but the scores for a given school system were highly unstable between the two periods. Predictions of teachers' mean descriptions of the superintendent and the superintendent's self-descriptions from the 1958 data to 1961 were no better than chance. Indeed, the data indicated that teachers and superintendents actively disagreed concerning the direction which the superintendents' behavior had moved in the three years.

A random half of the teachers in each school system of the replication study had received a modified version of the LBDQ schedule which we called the "personal" form. Most of the items in this form had been reworded so that they asked teachers to describe how the superintendent behaved toward them personally and individually rather than, as in the "regular" form, how he behaved toward "teachers in general." The "personal" form was devised on the grounds that it would be less subject to distortions by the teachers' own feelings and attitudes. The LBDQ scores were computed from these forms, too, and compared between the large and the small school systems. No statistically significant differences were found, not even on the Initiating Structure dimension which had revealed a difference (in the opposite direction from that of the Hunter finding) in the "regular" forms.

There were several minor ways in which the replication study did not correspond exactly with Hunter's investigation of 1958, and these were considered more closely for their possible bearing upon the lack of confirmation. The 20 school systems, for example, did not fall precisely in the same size ranges which Hunter had used, but when sub-groups of systems were selected which did, analyses still failed to reveal the expected differences in LBDQ scores associated with size. Another departure was a result of the investigator's attempt, in the replication study, to forestall a bias in the data due to the self-selection of respondents. Non-respondents in Hunter's study were replaced by additional subjects outside his original random sample; we sampled without substitutions. Data of the replication study, however, permitted us to reconstruct the non-random sample which Hunter had employed and to compare these scores for association with school system size. No association was found.

In short, the unusual pattern of findings in LBDQ scores which Hunter had uncovered in his dissertation study could not be detected in a study conducted under the most closely comparable circumstances. Why did they appear in the first place? We do not know. They were, after all, purely empirical findings--neither hypothesized in advance nor readily explainable as they stood--which emerged from one of the many implicit inspections he had made of his data. They obviously do not represent generalizable features of public school systems and their personnel; in fact, they cannot be generalized to the same school systems three years later. We can only declare them to be accidental findings whose further pursuit would be fruitless.

Reconsideration of the LBDQ in School Systems

With the disappearance of the initial problem of the study, the investigator's attention turned sharply on the LBDQ itself. What meaning do the scores hold? The conviction grew on us throughout our investigations that the LBDQ items told more about the feelings and attitudes of teachers, board members, superintendents, staff members--the respondents--than about how superintendents behaved as leaders. We systematically canvassed our own data and the findings from other studies of the leader behavior of school superintendents,[5] asking whether or not it is tenable to assume that LBDQ scores are simply veridical reflections of superintendent behavior plus random measurement error. That is, we asked for evidence of validity. Our canvass brought out the following points.

---

[5] Halpin, _op. cit._; Hunter, _op. cit._; Melvin Seeman, "A Comparison of General and Specific Leader Behavior Descriptions," in Stogdill and Coons, (Eds.) _op. cit._, pp. 86-102; and Melvin Seeman, "Social Mobility and Administrative Behavior," _American Sociological Review_, 1958, _23_, 633-642.

1. There is no substantive evidence, one way or another, from studies of school superintendents regarding concurrent, predictive, or construct validity of the LBDQ.

2. Respondents are not consistently capable of discriminating between the two supposedly independent dimensions of leader behavior. Dimension correlations are low in some samples but extremely high in other samples, even when both are reporting on the same superintendents.

3. Halo was a problem in the replication study, as it has been throughout the LBDQ's developmental history, but less so in the Initiating Structure dimension than the Consideration dimension and less in both dimensions of the "personal" form of the LBDQ.

4. Classes of respondents do not uniformly agree in the mean scores they furnish for their superintendents, and occasionally they actively disagree. The findings of Hunter's study which initiated our own explorations is a case in point. Generally speaking, absence of agreement is the rule.

5. Respondents within the same class do not agree appreciably among themselves. A special analysis conducted in the replication data demonstrated that teachers who had no first-hand knowledge about the superintendent's behavior were equally as likely to agree on how he behaved as members of the administrative staff who had the most favorable of vantage points for observing him.

6. The best that can be said is that the LBDQ instrument applied across a number of school systems is capable of generating scores which are comparatively unique for a given class of respondents in each school system. This observation can argue as strongly for the uniqueness of the respondents in each system as it can for the uniqueness of the superintendent's leader behavior.

While it would be foolhardy for us to propose that the superintendent's manner of behavior is completely unrelated to the reports which respondents furnish about it, it is apparent that objective facts of the behavior are not the dominating force underlying the reports. Perceptions of and reports on another person's behavior are intricate combinations of both the objective stimuli and the functional forces residing within the respondents themselves. Psychological theory instructs us that the optimal conditions under which a report reflects what is "out there" are those of high clarity and compellingness of the stimulus object and low strength of the reporter's attitudes, needs, moods, expectations and other functional forces. Conversely, optimal conditions for a report to protray what is inside the reporter himself are those of indeterminancy and ambiguity of the stimulus object and strong functional forces. It is difficult to sort out in detail the relative contributions of objective fact and of psychological processes to the LBDQ scores with the gross kind of data collected in studies such as our own, although an attempt was made in Chapter 8. This task would require investigations directed more intimately toward the psychology of the perceiver.

We are impressed, however, by the extent to which the use of the LBDQ, and particularly Hunter's revision of it, fulfills one of the conditions for obtaining information about the reporter himself. Hunter's one-word revision of the items, changing the word "staff" to "teachers," altered the context of the leadership act and rendered many of the items ambiguous. Content of the LBDQ items is predicated on the relation of leader to follower in the context of the small, intimate work group, but superintendents do not stand in this kind of relationship to classroom teachers. The central tasks of education are not pursued by small, unstructured work groups of teachers and certainly not with superintendents as group leaders. The closest approximation to leaders of small, ad hoc groups of teachers in the formally organized school system are building principals or other lower echelon supervisors, and even this is not a close approximation. Superintendents, as our data showed, rarely have more than fleeting contacts with the bulk of the teaching faculty, and this seems to be as true of the smaller school systems we studied as of the large ones. While the proportion of teachers with whom superintendents interacted regularly was higher in the small systems, this could be accounted for in large part by the happenstance of their geographical proximity to teachers and the small town acquaintance opportunities, not by a different kind of superintendent-teacher role. Consequently, Hunter's LBDQ items must have been unreal for the respondents. They were "biased" against both large and small system superintendents in that they referred to kinds of behavior which superintendents of schools could not be expected to display. And many of the teachers who were queried on his behavior had not the slightest grounds for judgment other than hearsay. Respondents had but two options: either to decline to respond to the LBDQ, as about a quarter of them did, or to invest their responses with personal fancies and group-shared attitudes. On rational grounds at least, Hunter's revision of the LBDQ had all the appearances of a projective test.

Beyond the limitation introduced by Hunter's revision of the LBDQ, an important question arises as to whether a concept and measure of leadership developed with the small group context in mind is applicable at all to the bureaucratic setting of the school. Many of the functions of leadership, which can only be supplied by the personal acts of some agent in the _ad hoc_ work group, are built into the organizational structure of the bureaucracy. Impersonal rules and regulations govern behavior; allocation of responsibility is specified by the division of labor; inducement to perform is carried in an incentive system and most effectively policed by groups of colleagues and the teacher, himself, by virtue of his internalization of professional norms; even the leadership function of considerateness, we believe, can become institutionalized in organizational arrangements. When leadership functions are served by impersonal mechanisms of organization, their consequences cannot, on logical grounds, be attributed to the behavioral acts of any given person, superintendent, principal, school board president, or whoever. Uniquely personal contributions of officials may be limited to the essentially residual, perhaps trivial, aspects of leadership or to the relatively rare occasions when officials attempt to re-arrange the bureaucratic structure.

A Note on Descriptions of Leader Behavior

It has not been our intent in this report to depreciate the efforts of students who have attempted to measure leader behavior of school superintendents or officials in other organizational settings. On the contrary, it is the significance of such efforts and the potentialities they hold which has induced us to review them so closely. When instruments like the LBDQ are constructed for use in organizations, we would hope that the lessons we have learned, or more correctly, the observations we have made, in the course of our own consideration of the problems are taken into account. These can be summarized as five points.

Leadership functions in the organization. However one may choose to define leadership, headship, administrative behavior, or similar concepts, they inevitably designate contributions to cooperative endeavor which are made by human agents. As we have pointed out, contributions are also made by impersonal mechanisms of organizational structure, and it is important not to confuse the two. Careful organizational analysis necessarily must precede the construction of behavioral description items if we are to avoid our culturally derived inclination to glorify the power of the individual human agent. Such an analysis would provide a balanced perspective and permit the personalistic descriptions to focus upon those functions which the human being is uniquely capable of serving.

The way teachers and other participants in the organization conceive of the personal contributions of officials to the various organization functions is a question apart, but an interesting one. We cannot assume, of course, that their conceptions coincide with a detached, logical analysis. The studies of Heider and others on the "psycho-logic" of perception certainly disavow that assumption. Specifically, Heider's analysis suggeststhat people differ in their inclination to attribute the cause of events to persons or to environmental sources--in their "propensity for heteronomous versus autonomous attribution" of causality.[6] Where two respondents may both believe that the superintendent's attitudes are "clear to teachers," (an LBDQ item) one may perceive the event as due to the intention and ability of the superintendent to make them clear while the other perceives it as due to the conditions in which the superintendent has found himself, irrespective of either his intent or his ability. Certainly, a deep understanding of leadership processes would entail investigations into the way participants weight their assessments of the behavior of officers in the face of environmental constraints.

---

[6] Fritz Heider, The Psychology of Interpersonal Relations, New York: John Wiley, 1958. P. 169.

The leader-follower relationship. As we saw in connection with Hunter's revision of the LBDQ, it is of more than casual interest to be clear about who is involved in the leader-follower relationship. Leader behavior is directed behavior--directed toward a specific class of other persons who constitute the followers. When Hunter redefined the leader follower dyad from "staff" to "classroom teahcer," the entire role system was altered and the LBDQ items, we believe, rendered inappropriate. The role relationship between superintendent and administrative staff is different from the relationship between the superintendent and teachers or school board members, and the circumstances in which the superintendent may act as a leader in the relationship varies correspondingly. The same set of behavioral description items cannot be used indiscriminately across all of the role sets in which the superintendent participates unless theitems have been carefully screened for their general applicability. It may well be that in a well structured organizational setting substantively different sets of items will have to be employed for the various kinds of relationship. In any event, the investigator is advised to consider the particular leader-follower relationship in which he is interested and how it differs from alternative relationships.

Kinds of information about behavior. The problem of describing another person's behavior, and of constructing questionnaire items for the purpose, is more complicated than it appears on the surface. Many of the difficulties are well explicated in the methodological literature on observation schedules. To cite an example, behavior may be described at many levels of generality from the most molecular to the most molar, and the investigator must decide where on the continuum his descriptions will fall. Are they to be molecular descriptions, requiring observations of immediate, concretistic muscle movements? Probably not, despite the high reliability with which such behaviors can be described, since they usually furnish little significant information. The investigator probably will wish to characterize more extended units of activities, sequences of the more molecular units which seem to him to represent consistent wholes. As the investigator directs his interests toward the more molar units of behavior, it seems that he must place greater reliance upon observer inference about the behavior in question. The price he pays for the more significant information is heterogeneity in the inferences which observors can draw from the same set of concrete actions. In the absence of well-conceived instructions and items to direct the inference process, observors can be expected to disagree with one another.

Heider's analysis is especially instructive in guiding the construction of items of behavioral description, for the process of responding to the items is centrally a process of interpersonal perception. His analysis suggests that the observor may choose to report on (1) intentions behind behavior, (2) the behavior itself, or (3) the consequences for others of the behavior.[7] The LBDQ items are a conglomerate of all, and the possibility resides in any given item for alternative perspectives to be applied. Neither the item wording nor the instructions specify which perspective the respondent should adopt, reducing thereby the liklihood that all will adopt the same one.

Access to information about behavior. The preceding two points bear intimately upon another question: Who, among persons privy to the leadership act, has the most direct access to information about the behavior? In other words, who are the most trustworthy respondents from the standpoint of being in a position to know? It is not necessarily the case that the person who authors the acts is in the best position to know. It will depend upon the kind of information sought. Superintendents, for example, can "observe" the intentions behind their behavior more directly than anyone else but often are in an unfavorable position to observe the consequences of their behavior for other persons in the leader-follower relationship. Those most remote from the superintendent may have no option but to report on consequences, while those closer can observe the behavior itself. Substantially greater knowledge is needed regarding systematic differences in the perspectives adopted by respondents, but it is not too soon to employ the fruits of analyses such as Heider's in planning and constructing behavioral description items. The investigator needs to ask whether his would-be respondents are in any position to furnish the kind of information which he seeks through his questionnaire items.

---

[7] Ibid. Especially Chapters 4 and 6.

Also involved in this question is the respondent's location with respect to the leader-follower relationship upon which he must report. In Halpin's Ohio study, for example, reports were obtained from staff members, school board members, and superintendents; his questionnaire items implicitly, and sometimes explicitly, inquired into the superintendent's behavior toward the staff, not toward the board. (It is quite possible that some of Halpin's respondents did not recognize this implication in the items where it was not made fully explicit and adopted the alternative point of view.) Thus, school board members stood as outside observors to the relationship, while superintendents and staff members were participating observors. The same thing occurred in Hunter's study, except that the leader-follower relationship was between superintendent and classroom teacher and the items left no room for equivocation on the point. In neither study was it likely that the board member had access to the sort of information he was asked to furnish. He could tell how the superintendent behaved in board meetings and, perhaps, infer from this how he might behave in other contexts, but it is obvious that a difficult demand was placed on him.

At another point in the replication study, we did anticipate the issue of accessibility of the reporter to the necessary information. We saw the possibility that the classroom teacher was in no position to describe the superintendent's behavior toward "teachers in general" and went on to construct the "personal" form of the LBDQ. We believed that teachers had the most dependable knowledge about how the superintendent behaved toward them individually. Our anticipation seemed to have been borne out by virtue, first, of the markedly different (lower) LBDQ scores on the "personal" form from those on the "regular" form and, second, of the smaller inclination of the "personal" forms to reflect halo.

In short, our experience suggests that it is critical for the investigator to assess carefully the opportunities for his would-be respondents to observe the kinds of behavior upon which they will be asked to report. This must be done as he constructs his questionnaire items. If he inadvertantly puts some classes of respondent at a disadvantage in providing the information his items seek, comparability of the responses across respondents will be impaired.

Behavior versus reactions to behavior. Finally, we must raise a question about the significance of halo in behavioral descriptions. Halo is a problem so long as the reports are considered to be measures of the objective fact of behavior. When that is his intent, the instrument maker will attempt to minimize the correlation between scores on the behavioral measure and measures of the respondents' generalized evaluations of or affect toward the person upon whom they are reporting.

From another point of view, however, halo is not a problem. One can argue that the presence of correlations between behavioral descriptions and affective measures are matters of indifference. Correlations by themselves furnish no clues concerning the direction of causality. It is just as likely that feelings about the leader are products of his behavior as it is that perceptions of his behavior are shaped by generalized feelings. One might ask why it is necessary to distinguish the direction of causality. Providing that the respondent is a participant in the leader-follower relationship and not an outside observor, the argument runs that the feelings about him are equally as significant as the facts of the leader's behavior and are equally as appropriate to measure.

The adoption of this point of view has its price. If behavior and affect are not separated, certain kinds of investigation and productive knowledge are foreclosed. It means, for example, that we will not be able to learn about the conditions under which different modes of leader behavior lead to different levels or kinds of affect among participants in the relationship. A leader would have no systematic basis for predicting the consequences in follower affect were he to change his mode of behavior. Which point of view the investigator chooses will depend upon his purposes, but it is essential that he be cognizant of the limitations of each.

If the student wishes to maintain a separation between behavior and affect and yet to avoid the worst aspects of the halo problem, one strategy is open to him. He can measure leader behavior in a way which is methodologically independent of his measure of follower affect. Specifically, this entails obtaining indicators of leader behavior from sources other than the persons who furnish information about their affective states. Followers who report their feelings about the leader should not be asked also to describe the leader. The relationships which are spuriously generated when this procedure is followed are fully documented in books on research methodology. Descriptions of leader behavior must be obtained either from dependable outside observors, by means of items appropriately designed, or through objective indicators unmediated by the inferences of respondents.

Appendix Tables

Table I

Mean LBDQ Scores from Teachers, by Teaching Level and School System Size: Hunter Data.*

System Size and Teaching Level	Initiating Structure	Consideration	(Number of Teachers)
Large			
Senior high school	36.0	39.1	(48)
Junior high school	35.4	39.6	(40)
Elementary school	35.4	39.7	(35)
Small			
Senior high school	41.9	46.5	(50)
Junior high school	41.7	46.2	(28)
Elementary school	42.2	47.2	(38)

---

*Computations made by the investigator from raw data furnished by Ovid N. Hunter.

Table II

Items in "Personal" Form of the LBDQ*

Instructions

Please indicate on the following 30 questions how you believe your superintendent behaves as a leader. Each item describes a specific kind of behavior without evoking any judgement regarding the desirability or undesirability of that behavior. These questions in no way constitute a "test" of the ability of the person who answers the items; nor do they involve an evaluation of the effectiveness of the superintendent's performance.

Please check all items.

The superintendent of my school . . .

1. Does personal favors for me.
2. Makes his attitudes clear to me.
3. Does little things to make it pleasant to be a teacher.
4. Tries out his new ideas with me.
5. Is easy for me to understand.
6. Rules me with an iron hand.
7. Finds time to listen to me.
8. Criticizes my poor work.
9. Speaks to me in a manner not to be questioned.
10. Keeps to himself where I am concerned.
11. Looks out for my personal welfare.
12. Assigns me to particular tasks.
13. Works without a plan.
14. Maintains definite standards of performance for me.
15. Refuses to explain his actions to me.
16. Acts without consulting me.
17. Is slow to accept new ideas from me.
18. Emphasizes my meeting deadlines.
19. Treats me as his equal.
20. Encourages the use of uniform procedure.
21. Is willing to make changes.
22. Makes sure that I understand his part in the organization.
23. Is friendly and approachable for me.
24. Asks that I follow standard rules and regulations.
25. Makes me feel at ease when talking with him.
26. Lets me know what is expected of me.
27. Puts my suggestions into operation.
28. Sees to it that I am working to capacity.
29. Gets my approval on important matters before proceeding.
30. Sees to it that my work is coordinated with others.

*Response categories on the right of the form were "always," "often," "occasionally," "rarely," and "never."

Table III

LBDQ Data by School Systems, from Teachers, Administrative Staff Members, and Superintendents: Replication Study

	SUPERINTENDENT		ADMINISTRATIVE STAFF					
			Initiating Structure			Consideration		
School System Code	I.S.	Cons.	Mean	Var.	(No.)	Mean	Var.	(No.)
Large								
00	37	50	39.4	80.30	(5)	42.2	169.70	(5)
01	49	55	39.8	14.20	(5)	49.6	34.80	(5)
02	41	45	40.1	47.81	(7)	45.6	82.62	(7)
03	33	44	38.0	17.20	(6)	53.5	18.70	(6)
04	31	42	42.7	69.33	(3)	43.3	190.33	(3)
05	30	45	40.7	5.07	(6)	45.0	63.00	(5)
06	31	45	46.4	81.50	(6)	48.4	37.87	(6)
07	45	49	35.6	16.29	(7)	42.7	22.95	(7)
08	34	36	37.6	21.95	(7)	46.9	12.48	(7)
09	33	39	32.4	81.29	(7)	40.4	43.95	(7)
Size Class*	36.4	45.0	39.3	49.86	(59)	45.8		(58)
Small								
10	42	48	36.2	58.97	(6)	40.2	37.77	(6)
11	31	35	32.8	8.33	(3)	40.8	7.00	(3)
12	34	41	29.3	60.92	(4)	45.0	97.58	(4)
13	45	47	32.2	3.07	(6)	43.5	14.17	(6)
14	35	38	37.5	41.58	(4)	49.5	63.00	(4)
15	45	41	40.3	46.33	(3)	38.7	134.33	(3)
16	47	49	29.5	16.67	(6)	38.8	27.07	(6)
17	31	46	48.7	12.50	(2)	52.8	12.50	(2)
18	32	45	30.7	10.33	(3)	32.8	74.92	(4)
19	-	-	14.7	213.50	(6)	26.7	186.97	(6)
Size Class*	38.0	43.3	34.6	114.06	(43)	41.0	101.00	(44)

---

*Means and variances computed over all personnel in size class. The mean of school system means and variances will differ from the figures shown here.

Table III (Cont.)

	TEACHERS							
	"Regular" Form				"Personal" Form			
	Initiating Structure		Consideration		Initiating Structure		Consideration	
School System Code	Mean	Var.	Mean	Var.	Mean	Var.	Mean	Var.
Large								
00	40.2	72.81	39.3	87.23	34.3	32.75	33.4	164.78
01	38.3	34.62	41.1	96.61	28.4	64.92	38.5	49.77
02	37.4	77.96	42.0	183.68	30.6	57.13	39.3	50.66
03	34.9	55.88	46.8	91.66	34.4	57.47	42.0	116.40
04	36.6	64.66	39.2	136.44	31.3	71.96	31.1	73.27
05	40.1	44.29	44.2	65.69	34.7	88.27	33.8	182.97
06	37.5	87.50	39.6	172.30	27.1	284.29	31.8	203.77
07	43.4	28.18	41.1	92.50	39.2	87.97	33.7	160.24
08	40.8	23.36	43.6	52.03	32.2	67.19	33.7	73.75
09	36.5	53.50	36.8	125.66	29.5	90.64	30.1	165.91
Size Class*	38.7	56.00	41.3	109.16	32.1	85.57	36.8	118.94
Small								
10	35.0	73.45	46.8	31.03	32.1	67.74	39.2	49.79
11	36.0	64.25	41.4	73.66	29.0	158.20	44.9	99.09
12	35.2	32.15	38.2	86.89	27.1	128.84	41.8	88.18
13	43.0	50.40	43.8	131.40	33.3	68.71	43.5	70.40
14	38.4	87.46	48.8	50.93	33.3	73.52	43.5	60.94
15	31.0	74.54	33.5	234.77	26.5	115.44	25.5	345.60
16	39.2	45.57	37.1	212.26	36.0	56.69	33.1	203.49
17	28.1	30.38	38.2	53.41	27.8	173.76	38.4	107.05
18	34.8	72.90	29.2	160.77	28.2	94.19	30.5	91.91
19	29.7	75.70	32.4	146.42	21.3	103.79	37.4	93.87
Size Class*	35.1	75.94	38.7	226.29	29.8	109.32	37.6	148.50

---

*Means and variances computed over all teachers in size class. The means of school system means and variances will differ from the figures shown here.

Table IV

Items of the Teacher Affect Scales

Item No.	P Scale
5	The longer I know my superintendent the better I like him.
8	Most of the people whose opinions I respect feel our superintendent is doing more harm than good for the school.
12	My superintendent is not the kind of person I normally would seek as a friend.
16	Few of the teachers I know speak with pride of our superintendent.
20	Many teachers in the system would be relieved if our superintendent took another job.
23	Taking everything into account, the superintendent in my district is better than the ones in most districts around here.
	A Scale
7	It is unfair to say that teacherscontribute more to effective schools than top level administrators.
10	It is a waste of good talent to move a top-notch building principal into a higher-level administrative job in the system.
13	If I had to make a choice, I would say a well-qualified superintendent contributes more to good education than a well-qualified building principal.
21	One good way to improve schools in this country would be to give superintendents more administrative help.
24	The important work in education is done by the teachers, not by the staff members around the superintendent's office.
	S Scale
2	I am proud to say I teach in this school system.
3	I would certainly encourage a beginning teacher to take a job in this district.
9	Parents in thiscommunity have good reason to brag about their schools.
15	Most of the teachers I know here consider themselves lucky to be teaching in a school system such as ours.
18	I would rather teach in most any other school system than this one.
19	If I were a parent of school-age children, I would think twice before moving into this school district.
22	I would be sorry if I had to move from this district to take a job in another one.

Table IV (Cont.)

Item No.	T Scale
1	The youngsters of today are a joy to teach.
4	I have gotten so I dread meeting my classes in the morning.
6	For me classroom teaching is a very satisfying type of work.
11	I look forward to a better kind of life than teaching school.
14	The personal rewards I get from classroom teaching outweigh the trials and tribulations it so often involves.
17	Every year in teaching is another year of my life wasted.

Table V

Details of Scale Analyses Conducted on Trial Run Data

P SCALE

Response Pattern

	Item Number*						Scale Type	
	16	8	5	23	20	12	Perfect	Total
	+	+	+	+	+	+	12	22
	-	+	+	+	+	+	11	12
	-	-	+	+	+	+	10	15
	-	-	-	+	+	+	5	10
	-	-	-	-	+	+	5	12
	-	-	-	-	-	+	7	8
	-	-	-	-	-	-	13	16
Sum							63	95
Errors	4	6	7	11	4	6 = 38		
Marginal	24%	36%	52%	62%	75%	83%		

Per Cent Perfect = 66%
Reproducibility =.93

Item Scoring

Item Number*	SA	A	?	D	SD
16	-	-	-	-	+
8	-	-	-	-	+
5	+	+	-	-	-
23	+	+	-	-	-
20	-	-	-	+	+
12	-	-	+	+	+

---

*Numbered according to final questionnaire form.

Table V (Cont.)

A SCALE

Response Pattern

	Item Number*						Scale Type	
	24	13	10	7**	21		Perfect	Total
	+	+	+	+			6	19
	-	+	+	+			12	22
	-	0	+	+			10	12
	-	-	+	+			15	16
	-	-	-	+			11	13
	-	-	-	-			4	9
Sum							58	91
Errors	5	11	13	5		=34		
Marginal	26%	42%	53%	84%				

Per cent Perfect = 64%
Reproducibility =.93

Item Scoring

Item Number*	SA	A	?	D	SD
24	-	-	+	+	+
13	+	+	0	-	-
10	-	-	+	+	+
7	+	+	0	-	-
21	+	+	+	-	-

---

*Numbered according to final questionnaire form.
**Contrived item. Scored - for any combination of - in Item 21 and - or 0 in Item 7. Scored + for any other combination.

Table V (Cont.)

S SCALE

Response Pattern

	Item Number*								Scale Type	
	22	2	19	18	9	15	3		Perfect	Total
	+	+	+	+	+	+	+		8	16
	-	+	+	+	+	+	+		4	8
	-	-	+	+	+	+	+		7	9
	-	-	-	+	+	+	+		5	10
	-	-	-	-	+	+	+		12	19
	-	-	-	-	-	+	+		4	7
	-	-	-	-	-	-	+		4	8
	-	-	-	-	-	-	-		8	15
Sum									52	92
Error	2	11	6	3	10	10	8	= 50		
Marginal	21%	29%	39%	50%	72%	80%	83%			

Per cent perfect = 57%
Reproducibility =.92

Item Scoring

Item Number*	SA	A	?	D	SD
22	+	-	-	-	-
2	+	-	-	-	-
19	-	-	-	-	+
18	-	-	-	-	+
9	+	+	-	-	-
15	+	+	+	-	-
3	+	+	+	-	-

---

*Numbered according to final questionnaire form.

Table V (Cont.)

T SCALE

Response Pattern

	Item Number*						Scale Type	
	11	14	6	4	17	1	Perfect	Total
	+	+	+	+	+	+	16	24
	-	+	+	+	+	+	5	10
	-	-	+	+	+	+	5	13
	-	-	-	+	+	+	2	8
	-	-	-	-	+	+	6	6
	-	-	-	-	-	+	17	19
	-	-	-	-	-	-	11	14
							62	94
Sum Errors	6	10	7	9	8	3 = 43		
Marginal	35%	37%	48%	55%	64%	68%		

Per cent perfect = 66%
Reproducibility =.92

Item Scoring

Item Number*	SA	A	?	D	SD
11	-	-	-	-	+
14	+	-	-	-	-
6	+	-	-	-	-
4	-	-	-	-	+
17	-	-	-	-	+
1	+	+	-	-	-

---

*Numbered according to final questionnaire form.

Table VI

Factor Loadings from Oblique Rotation of Principal Axis Factor Analysis of Thirty Items in Five Cumulative Scales: Trial Run Data

	Factors					
Item No.	I	II	III	IV	V	VI
5	59	08	-03	04	05	06
8	-60	01	-01	03	-01	14
12	-63	-04	16	-05	-04	-02
16	-54	02	-02	06	-05	12
20	-67	01	-04	11	05	14
23	53	01	05	-01	-13	27
7	04	71	04	-03	-05	09
10	-06	-31	01	-04	-06	44
13	02	57	04	10	-13	36
21	08	26	-07	16	15	11
24	01	-63	00	16	-10	38
2	-08	00	67	07	01	-08
3	06	08	58	-12	04	-09
9	06	05	64	02	00	-04
15	01	04	59	-06	04	00
18	-13	10	-48	01	10	41
19	04	08	-60	10	00	41
22	01	07	53	18	-13	04
1	-28	08	12	57	16	-12
4	-15	10	-01	-48	04	15
6	-09	-10	-03	75	06	14
11	-04	03	04	-64	32	05
14	-01	-06	-01	69	00	-02
17	-19	02	01	-45	13	40
a	11	-12	11	01	12	24
b	-20	02	-04	03	-45	-09
c	-16	08	-03	-02	-48	-05
d	12	-03	10	00	-67	-04
e	-07	-06	10	07	55	12
f	-03	01	-02	-03	-59	04

Table VII

Teacher Reaction Questionnaire

If you are no longer teaching
or if you have changed school
systems since the beginning
of school, please check here
and return the questionnaire
to us _____.

Section A

Instructions

First, we need to know a number of things about you as a teacher. Please fill in the blanks or check the appropriate boxes below.

1. Male _____
   Female _____

2. Age. Please write in your age to your nearest birthday _____.

3. Teaching level.
   Elementary _____
   Junior High _____
   Senior High _____

4. Experience. Number of years of teaching experience (counting this year as a full year)_____.

5. Tenure in system. How many years have you been teaching continuously in your present school system (counting this year as a full year)? ____.

6. Have you ever taught in a school system other than your present one?
   Yes ____
   No ____
   If yes, how long ago was it ___?

7. About how maby classroom teachers are in the building in which you teach? _____

8. Considering only the last three years, have you . . .

   a. Served on any official committees in your building?
      Yes ____
      No ____

   b. Served on any system-wide committees on which teachers from other buildings or persons from the central administration were also members?
      Yes ____
      No ____
      If "yes," were you chairman any of the time?
      Yes ____
      No ____

   c. Have you held a job in the school or system other than classroom teacher, either part or full time, e.g., principal, supervisor, head teacher, department or grade-level chairman?
      Yes ____
      No ____
      If "yes," what was it? ___.

d. Had any responsibilities in your school or school system beyond that of classroom teacher, e.g., in charge of audiovisual aids in the building?
   Yes ___
   No ___
   If "yes," what? ___

9. Were you graduated from a high school in the district in which you are now teaching?
   Yes ___
   No ___

10. What kind of work did your father (or head of household) do while you were growing up?
   Professional ___
   Business, managerial ___
   Clerical or white collar ___
   Skilled labor ___
   Other labor ___
   Farm ___

11. Considering the members of the school board and the two or three administrators who are at the top of the system

a. Are you on a first-name basis with any of them?
   Yes ___
   No ___

b. Have any of them visited in your home?
   Yes ___
   No ___

c. Are any of them closely related to you?
   Yes ___
   No ___

d. Are any of them members of a club, lodge, or other social organization to which you belong?
   Yes ___
   No ___

12. Over the last two or three years about how often have you seen the superintendent of your school to talk with him face-to-face? (Check one only)
   Several times a week ___
   Several times a month ___
   Every one or two months. ___
   Several times a year. ___
   Not more than once or twice a year. ___
   Have never talked with him in the last two or three years. ___

13. If you have talked with your superintendent in the last two or three years, which of the following describes the nature of your contacts? (Check all phrases that apply)
   Our contacts were limited to brief casual remarks on the events of the day. ___
   We have had more extensive conferences on school business. ___
   I have discussed my personal or professional problems with him. ___
   He and I have talked as friends about our common personal interests and affairs. ___

14. Suppose your school board or the superintendent's office were to announce a change in procedures, e.g., lengthening the school day or school year, drastically reducing sick leave or something similar which was highly unpopular with the teachers. Is there any local group or organization among the teachers which could both make clear the teachers' viewpoint and eventually cause the unpopular decision to be changed?
   Yes ___
   No ___
   If "yes," what group or organization is it? ___

Section B*

Instructions

Please indicate on the following 30 questions how you believe your superintendent behaves as a leader. Each item describes a specific kind of behavior without evoking any judgment regarding the desirability or undesirability of that behavior. These questions in no way constitute a "test" of the ability of the person who answers the items; nor do they involve an evaluation of the effectiveness of the superintendent's performance. (Please check all items).**

The superintendent of my school . . .

1. Does personal favors for me.
2. Makes his attitudes clear to me.
3. Does little things to make it pleasant to be a teacher.
4. Tries out his new ideas with me.
5. Is easy for me to understand.
6. Rules me with an iron hand.
7. Finds time to listen to me.
8. Criticizes my poor work.
9. Speaks to me in a manner not to be questioned.
10. Keeps to himself where I am concerned.
11. Looks out for my personal welfare.
12. Assigns me to particular tasks.
13. Works without a plan.
14. Maintains definite standards of performance for me.
15. Refuses to explain his actions to me.
16. Acts without consulting me.
17. Is slow to accept new ideas from me.
18. Emphasizes my meeting deadlines.
19. Treats me as his equal.
20. Encourages the use of uniform procedures.
21. Is willing to make changes.
22. Makes sure that I understand his part in the organization.
23. Is friendly and approachable for me.
24. Asks that I follow standard rules and regulations.
25. Makes me feel at ease when talking with him.
26. Lets me know what is expected of me.
27. Puts my suggestions into operation.
28. Sees to it that I am working to capacity.
29. Gets my approval on important matters before proceeding.
30. Sees to it that my work is coordinated with others.

How would you estimate the over-all leadership behavior of your superintendent? Circle the number below showing your estimate of his quality as a leader.

1 2 3 4 5 6 7
Poor Perfect

---

*See Table 4, p. 19 for wording of items in the "Regular" form.

**Response alternatives were "Always," "Often," "Occasionally," "Rarely," and "Never."

Section C

In the following items you are asked to react to certain aspects of your teaching situation. As before, there is no "right" or "wrong" answer and the information you give will not be used in any evaluative sense. (Please score all items)*

Part 1

In the school system where I work, a teacher like myself . . .

a. Believes he has some control over what textbooks will be used in the classrooms.
b. Feels he does not know what is going on in the upper levels of administration.
c. Never has a chance to work on school committees which make important decisions in the school system.
d. Considers that he has little to say over what teachers will work with him on his job.
e. Usually can find ways to get system-wide policies changed if he feels strongly enough about them.
f. Feels he has little to say about important system-wide policies relating to teaching.

Part 2

1. The youngsters of today are a joy to teach.
2. I am proud to say I teach in this school system.
3. I would certainly encourage a beginning teacher to take a job in this district.
4. I have gotten so I dread meeting my classes in the morning.
5. The longer I know my superintendent the better I like him.
6. For me classroom teaching is a very satisfying type of work.
7. It is unfair to say that teachers contribute more to effective schools than top level administrators.
8. Most of the people whose opinions I respect feel our superintendent is doing more harm than good for the school.
9. Parents in this community have good reason to brag about their schools.
10. It is a waste of good talent to move a top-notch building principal into a higher-level administrative job in the system.
11. I look forward to a better kind of life than teaching school.
12. My superintendent is not the kind of person I normally would seek as a friend.
13. If I had to make a choice, I would say a well-qualified superintendent contributes more to good education than a well-qualified building principal.
14. The personal rewards I get from classroom teaching outweigh the trials and tribulations it so often involves.

---

*Response alternatives were "Strongly agree," "Agree," "Maybe and maybe not," "Disagree," and "Strongly disagree."

15. Most of the teachers I know here consider themselves lucky to be teaching in a school system such as ours.
16. Few of the teachers I know speak with pride of our superintendent.
17. Every year in teaching is another year of my life wasted.
18. I would rather teach in most any other school system than this one.
19. If I were a parent of school-age children, I would think twice before moving into this school district.
20. Many teachers in the system would be relieved if our superintendent took another job.
21. One good way to improve schools in this country would be to give superintendents more administrative help.
22. I would be sorry if I had to move from this district to take a job in another one.
23. Taking everything into account, the superintendent in my district is better than the ones in most districts around here.
24. The important work in education is done by the teachers, not by the staff members around the superintendent's office.